

**UNIVERSIDAD AUTÓNOMA DE MADRID**  
**DEPARTAMENTO DE BIOQUÍMICA**



**Caracterización de variantes genéticas en regiones reguladoras  
que afectan a la respuesta a hipoxia y su papel en enfermedades  
respiratorias obstructivas crónicas**

**Ma. Laura Deguiz Lasso**

**Madrid, 2015**



**DEPARTAMENTO DE BIOQUÍMICA  
FACULTAD DE MEDICINA  
UNIVERSIDAD AUTÓNOMA DE MADRID**

**Caracterización de variantes genéticas en regiones reguladoras  
que afectan a la respuesta a hipoxia y su papel en enfermedades  
respiratorias obstructivas crónicas**

**Ma. Laura Deguiz Lasso  
Licenciada en Ciencias Biológicas**

**Directores de tesis:  
Benilde Jiménez Cuenca  
Luis del Peso Ovalle**

**Instituto de Investigaciones Biomédicas “Alberto Sols”, Madrid  
Consejo Superior de Investigaciones Científicas – Universidad Autónoma de Madrid**



Benilde Jiménez Cuenca y Luis del Peso Ovalle, Profesores Titulares del Departamento de Bioquímica de la Facultad de Medicina de la Universidad Autónoma de Madrid.

CERTIFICAN: que Dña. María Laura Deguiz Lasso, Licenciada en Ciencias Biológicas por la Universidad de Buenos Aires, ha realizado bajo nuestra dirección el trabajo de investigación titulado:

Caracterización de variantes genéticas en regiones reguladoras que afectan a la respuesta a hipoxia y su papel en enfermedades respiratorias obstructivas crónicas

Y que éste reúne todas las condiciones requeridas por la legislación vigente y la originalidad y calidad científica necesarias para poder ser presentado y defendido con el fin de optar al grado de Doctor.

Y para que conste donde proceda, firmo el presente certificado.

Madrid, 6 de Octubre de 2015

Benilde Jiménez Cuenca

Luis del Peso Ovalle





## AGRADECIMIENTOS

En primer lugar quiero agradecer a la Dra. Benilde Jiménez y el Dr. Luis del Peso, por permitirme ser parte de sus laboratorios y darme la oportunidad de continuar con mi formación profesional. Gracias por enseñarme a hacer ciencia e inculcarme que ello incluye no solo hacer experimentos sino también parar y poner la cabeza en ellos. El apoyo que me han brindado y su cercanía han hecho que las malas épocas hayan sido más fáciles de superar, muchas gracias a los dos.

Quiero continuar dando las gracias a todos los colaboradores que han hecho que esta tesis sea posible, principalmente a los Servicios de Neumología del Hospital Universitario La Paz y Hospital Universitario La Princesa, coordinados por el Dr. Francisco García Río y el Dr. Julio Ancochea, que se han encargado del aspecto clínico de este trabajo. También quiero agradecer a Olga quien ha comenzado este proyecto y con quien he compartido muchísimas horas, su ayuda y compañía han sido impagables, gracias de corazón.

Y nada de esto hubiese sido posible sin mis amigas y compañeras de trabajo, las que siguen aquí: Mery, Olga y Asun, y las que ya no están en el IIB pero siguen presentes: Sara, Laura y Bea. Quienes me han sacado de un apuro y dado consejos en más de una oportunidad. Pero también han sido muy importantes a nivel personal pues me recibieron desde el minuto cero como una más, haciéndome sentir como en casa, aguantándome en las buenas y en las malas. Siempre dispuestas a escucharme y darme un abrazo cuando lo necesitaba. Gracias por estar ahí siempre!

No me quiero olvidar la gente que forma parte del IIB con quien me cruzo en los pasillos, cultivos, comidas y, en los últimos meses, a las pretésicas de la biblioteca que me han escuchado reír y llorar, todos ustedes hacen que pasar mil horas en el laboratorio no sea una tortura.

Fuera del IIB, es mucha la gente que me ha apoyado y dado palabras de aliento cuando una luciferasa me arruinaba el día, por eso gracias infinitas a los guapitos de Espronceda (Kiko, Camilo, Flo, Morgan y Ana) que me han abierto las puertas de sus casas compartiendo su familia, sus amigos y sus días. A Barbi y Gri, gracias por estar siempre y aguantarme, es una alegría poder contar con ustedes. Y a Merche e Irene, con quienes siempre da gusto compartir un rato juntas.

Y más allá de Madrid, en Buenos Aires están todos los demás que hacen que mi vida sea completa: mis amigos y mi familia. La prueba de que eso de estar cerca no tiene nada que ver con la distancia porque los he sentido siempre a mi lado. Gracias!!!! A mis amigos de siempre (Nati, Gi, Eze y Fede), a los que me regalo la carrera (Cyn, Juli, Ana, Xime y Juan) y a la Familia del Gallo-Segura, GRACIAS POR ESTAR. A mis padres, mi hermano y mis abuelos, que siempre me apoyan y quienes han sido mi ejemplo de trabajo y esfuerzo. Nunca me voy a cansar de darles las gracias! Y a vos Diego por ser mi mejor compañía y mi mejor amigo, gracias por comprenderme siempre y darme un abrazo siempre que lo necesito. Muchas gracias.

La realización de esta Tesis Doctoral ha estado financiada por una beca de Formación de Personal Investigador y por los proyectos SAF-2010-19256 y SAF2011-24225 del Ministerio de Economía y Competitividad.



## RESUMEN

Las variantes genéticas en regiones no codificantes del genoma pueden alterar la regulación de la expresión génica y, como consecuencia, tener múltiples efectos funcionales. De acuerdo con esta posibilidad, resultados previos del laboratorio demostraron que un polimorfismo de un único nucleótido (SNP) en el sitio de unión del Factor Inducible por Hipoxia (HIF) puede alterar la inducción de genes específicos de hipoxia. Por otro lado, en enfermedades respiratorias, como la Enfermedad Pulmonar Obstructiva Crónica (EPOC) y el Síndrome de Apnea Obstructiva del Sueño (SAOS) la respuesta celular a hipoxia y por tanto la expresión génica regulada por HIF se encuentra activada en estos pacientes. Estas enfermedades tienen un origen multifactorial y presentan una evolución y pronóstico variable en función de la presencia de comorbilidades. Las causas de la evolución variable, incluida la aparición de comorbilidades, son desconocidas pero podría deberse a la diferente predisposición genética.

En este trabajo estudiamos la variación genética en sitios de unión de HIF y su asociación con los diferentes fenotipos y comorbilidades en pacientes de EPOC y SAOS. Para ello, integramos datos experimentales de sitios de unión de HIF identificando 283 regiones del genoma que unen HIF.

Posteriormente, estudiamos mediante secuenciación masiva de segunda generación (NGS) estas 283 regiones en muestras genómicas de 96 pacientes con EPOC y 144 pacientes con SAOS. Este análisis nos permitió identificar 72 variantes genéticas que alteran el núcleo central de unión de HIF, el motivo RCGTG. Para estudiar el efecto funcional de estas variantes, seleccionamos el SNP rs150921338 que se localiza en el promotor de *PGK1* y el SNP rs6593210 en una región intrónica de *EGFR* y analizamos su efecto en la expresión de estos genes. Nuestros resultados, demuestran que estas variantes suprimen la inducción génica de estos genes por hipoxia. Además, mediante la edición del genoma con el sistema CRISPR/Cas9, pudimos confirmar que el SNP rs6593210 destruye un HRE funcional que controla la inducción de *EGFR* por hipoxia en la línea celular HEK293T.

Por último, estudiamos la asociación entre las variantes genéticas identificadas y los diferentes fenotipos y comorbilidades de las enfermedades estudiadas. Nuestros resultados, pendientes de validación en una cohorte de pacientes independiente, sugieren que la variante genética rs118151281 se asocia con una mayor probabilidad de padecer hipertensión arterial y riesgo cardiovascular en pacientes de EPOC; y que el SNP rs1730190 y rs1657290 se asocia con una menor presencia de hipertensión arterial en pacientes de SAOS.

En conjunto estos resultados sugieren que la variabilidad genética en motivos RCGTG en regiones no codificantes afecta la respuesta génica a la hipoxia y podría contribuir a la heterogeneidad observada en la evolución de enfermedades respiratorias crónicas.



## ABSTRACT

Genetic variations in non-coding regions can modify gene expression and, as a consequence, lead to multiple functional effects. Previous results of our laboratory showed that single nucleotide polymorphism (SNP) in hypoxia inducible factor (HIF) binding sites could disrupt transcriptional gene induction in hypoxia.

On the other hand, it is known that in respiratory diseases, such as Chronic Obstructive Pulmonary Disease (COPD) and Obstructive Sleep Apnea Syndrome (OSAS), hypoxic cellular response is activated and gene regulation mediated by HIF is triggered in these patients. These diseases have a multifaceted nature and show a variable prognosis and evolution according to the comorbidities. The causes of such heterogeneity, including the development of comorbidities, are unknown but could be due to different genetic background.

In this study, we examined genetic variation in HIF binding sites and its association with phenotypes and comorbidities in COPD and OSAS patients. To do this, we integrated experimental data of HIF binding sites identifying 238 HIF binding regions in the genome.

Subsequently, we studied by massive second-generation sequencing (NGS) these 238 regions in genomic samples of 96 COPD and 144 OSAS patients. This analysis allowed us to identify 72 genetic variants modifying the central core of HIF binding site, RCGTG motif. To study the functional effect of these variants, we selected SNP rs150921338 that localize in the promoter region of *PGK1* and SNP rs6593210 in an intronic region of *EGFR* and we analyzed their effect the expression of these genes. Our results show that these genetic variants preclude gene induction in hypoxia. We confirmed through gene editing by CRISPR/Cas9 system, that SNP rs6593210 destroys the functional HRE that controls *EGFR* expression in hypoxia in HEK-239T cell line.

On the other hand, we studied associations between identified genetic variants and phenotypes or comorbidities of the studied diseases. Our results, pending validation in a separate cohort of patients, suggest that the variant rs118151281 associates with a higher risk of arterial hypertension and cardiovascular risk in COPD patients, and SNP rs1730190 y rs1657290 associates with a lower risk of arterial hypertension in OSAS patients.

Altogether, these results suggest that genetic variation in RCGTG motifs in non-coding regions could affect genetic response to hypoxia and could contribute to the observed heterogeneity in the evolution of chronic respiratory diseases.



## INDICE

<b>1</b>	<b>ABREVIATURAS.....</b>	<b>1</b>
<b>2</b>	<b>INTRODUCCIÓN .....</b>	<b>5</b>
2.1	<b>Homeostasis del oxígeno.....</b>	<b>7</b>
2.1.1	Estructura de HIF.....	8
2.1.2	Regulación de la actividad transcripcional de HIF. ....	10
2.1.3	Regulación de la expresión génica por HIF. ....	12
2.2	<b>Enfermedades respiratorias .....</b>	<b>14</b>
2.2.1	Enfermedad pulmonar obstructiva crónica (EPOC) .....	15
2.2.2	Síndrome de apnea obstructiva del sueño (SAOS) .....	18
2.2.3	Papel de HIF en EPOC y SAOS. ....	20
2.3	<b>Papel de las variantes genéticas en regiones reguladoras en la variabilidad fenotípica interindividuo.....</b>	<b>21</b>
<b>3</b>	<b>OBJETIVOS.....</b>	<b>25</b>
<b>4</b>	<b>MATERIALES Y MÉTODOS.....</b>	<b>29</b>
4.1	Cultivos celulares y tratamiento de hipoxia .....	31
4.2	Generación de plásmidos .....	31
4.3	Ensayos reporteros.....	32
4.4	Selección de pacientes .....	32
4.5	Extracción de ARN y qPCR-RT .....	32
4.6	Perfil de expresión génica de alta resolución.....	36
4.7	Análisis de proteínas.....	37
i)	Extracción de proteínas celulares.....	37
ii)	Inmunodetección de proteínas mediante Western blot .....	37
4.8	Determinación de regiones reguladoras que unen HIF en el genoma humano .....	38
4.9	Secuenciación masiva .....	40
4.10	Edición del genoma mediante el sistema CRISPR/Cas9 .....	42
i)	CRISPR/Cas9: Generación de plásmidos. ....	44
ii)	CRISPR/Cas9: Condiciones de transfección.....	45
iii)	CRISPR/Cas9: Ensayo Surveyor .....	45
iv)	CRISPR: Edición del genoma .....	47
4.11	Análisis de asociación entre SNPs identificados y los fenotipos de EPOC y SAOS. ....	47
4.12	Programas informáticos y análisis estadístico de los datos .....	49
<b>5</b>	<b>RESULTADOS .....</b>	<b>51</b>
5.1	<b>Identificación de regiones que contienen sitios de unión de HIF en el genoma.....</b>	<b>53</b>
5.1.1	Caracterización funcional de los estados de la cromatina en regiones reguladoras con sitios de unión de HIF. ....	55
5.1.2	Estudio de ontología de genes asociados a las regiones reguladoras con sitios de unión de HIF.....	57
5.2	<b>Identificación de variantes genéticas en sitios de unión a HIF, en pacientes con EPOC y SAOS. ....</b>	<b>58</b>
5.2.1	Identificación de variantes de un único nucleótido en sitios de unión de HIF por secuenciación masiva.....	59
5.2.2	Validación de variantes genéticas identificadas en motivos RCGTG. ....	64
5.3	<b>Caracterización funcional de SNPs en sitios de unión de HIF. ....</b>	<b>66</b>

5.3.1	Estudio funcional del efecto de SNPs en motivos RCGTG mediante ensayos reporteros.....	67
5.3.2	Análisis funcional del efecto de SNPs mediante edición del genoma en motivos RCGTG utilizando el sistema CRISPR/Cas9 .....	74
	Elección de ARN guía para edición de motivos RCGTG con el sistema CRISPR/Cas9. ....	75
	Edición de motivos RCGTG con el sistema CRISPR/Cas9. ....	78
5.4	<b>Correlación entre las variantes genéticas en sitios de unión de HIF y los fenotipos de la EPOC y el SAOS. ....</b>	<b>80</b>
6	<b>DISCUSIÓN .....</b>	<b>87</b>
6.1	Identificación de variantes genéticas que modifican sitios de unión HIF en el genoma. ....	89
6.2	Caracterización funcional de SNPs en sitios de unión de HIF.....	91
6.3	Edición del genoma como herramienta para el análisis funcional de las variantes genéticas en regiones reguladoras.....	93
6.4	Correlación entre las variantes genéticas en sitios de unión de HIF y los fenotipos de la EPOC y el SAOS. ....	95
7	<b>CONCLUSIONES.....</b>	<b>99</b>
8	<b>BIBLIOGRAFÍA.....</b>	<b>103</b>

## **1 ABREVIATURAS**





**A**

$\alpha$ 1 ATD: Alpha-1 Antitrypsin  
 ANOVA: análisis de la varianza con un factor  
 AP-1: activator protein 1  
 ATF1: activating transcription factor 1  
 ATP9A: ATPase, Class II, Type 9A

**B**

BCKDHA: Branched Chain Keto Acid Dehydrogenase E1, Alpha Polypeptide  
 bHLH: hélice-bucle-hélice  
 BODE/BODEx: Body mass index, airflow Obstruction, Dysnea and Exercise capacity index.  
 BNIP-3: BCL2/adenovirus E1B 19 kDa protein-interacting protein 3.

**C**

Cas9: CRISPR associated protein 9  
 ChIP: Chromatin Immunoprecipitation. Inmunoprecipitación de cromatina.  
 CNS: Secuencias no codificantes conservadas evolutivamente.  
 CO<sub>2</sub>: Dióxido de carbono  
 COTE: CO-morbidity Test. Test de comorbilidad.  
 COPD: Chronic Obstructive Pulmonary Disease  
 CREB1: CAMP responsive element binding protein 1.  
 CRISPR: clustered regularly interspaced short palindromic repeats  
 Ct: umbral de detección establecido.  
 CTAD: Dominio de transactivación carboxilo terminal.  
 CTCF: 11-zinc finger protein  
 CTD1: carboxy-terminal domain, RNA polymerase II, polypeptide A, phosphatase, subunit 1

**D**

D/A: Depresión/ Ansiedad  
 DAG: Diacilglicerol  
 DL: Dislipidemia  
 DM: Diabetes mellitus  
 DMOG: Dimethyloxallyl Glycine  
 DNAJB6: DnaJ (Hsp40) Homolog, Subfamily B, Member 6  
 DPT: Dermatopontin

**E**

ECE2: Endothelin Converting Enzyme  
 EGFR: Epidermal growth factor. Receptor del factor de crecimiento epidérmico  
 EGLN3: Egl nine homolog 3.  
 EGLNs: Egl nine homolog. Homólogo de Egl-nine C. Elegans.

EHD2: EH-Domain Containing 2

EMSA: Electrophoretic mobility shift assay.  
 EPAS1: Endothelial PAS domain-containing protein 1  
 EPO: Eritropoyetina.  
 EPOC: Enfermedad pulmonar obstructiva crónica  
 ERAM: Esfuerzos respiratorios asociados a microdespertares  
 ESD: excesiva somnolencia diurna

**F**

FAM162A: Family with Sequence Similarity 162, Member A  
 FAM57A: Family with Sequence Similarity 57, Member A  
 FEV1: Volumen Espiratorio Máximo en el primer Segundo.  
 FGF11: Fibroblast Growth Factor 11  
 FUT11: Fucosyltransferase 11  
 FVC: Capacidad vital forzada  
 FDR: False Discovery Rate.  
 FIH: Hypoxia-inducible factor 1-alpha inhibitor. Factor Inhibidor de HIF.

**G**

GEO: Gene Expression Omnibus  
 GOLD: Global Initiative for Chronic Obstructive Lung Disease  
 GTF2IRD2B: GTF2I Repeat Domain Containing 2B

**H**

H3K27Ac: Histone H3 (acetyl K27)  
 H3K27me3: Histone H3 tri methyl K27  
 H3K36me3: Histone H3 tri methyl K36  
 H3K4me: Histone H3 methylation K4 H3K4me2: Histone H3 di methyl K4  
 H3K4me3: Histone H3 trimethyl Lys4  
 H3K9ac: Histone H3 acetyl K9  
 H4K20me: Histone H4 lysine 20 methylation  
 HBS: *HIF binding site*. Sitio de unión de HIF.  
 HDR: Homology directed repair. Recombinación homóloga.  
 HIF: *Hypoxia Inducible Factor*. Factor Inducible por Hipoxia  
 HILDPA: Hypoxia-inducible Lipid Droplet-associated  
 HMOX1: Heme oxygenase-1  
 HRE: Hypoxia responsive element. Elemento de respuesta a hipoxia.  
 HTA: Hipertensión Arterial

**I**

IAH: índice de apneas/hipopneas

IAR: índice de alteración respiratoria  
 IGFBP3: insulin-like growth factor binding protein 3  
 IGV: Integrative Genomics Viewer  
 IMC: índice de masa corporal  
 INDELS: inserciones/delecciones  
 INSIG1: Insulin-induced gene 1 protein  
 IP3: inositol 3 fosfato  
 IL-8: Interleucina 8.

## K

KDM4B: Lysine (K)-Specific Demethylase 4B

## L

LDHA: *L-lactate dehydrogenase A chain*.  
 LOX: Lysyl Oxidase

## M

MAB21L3: mab-21-like 3  
 MIF: macrophage migration inhibitory factor  
 MMRC: Modified Medical Research Council dyspnea Scale  
 m-TOR: mammalian Target of Rapamycin  
 MUT-HRE: variantes que afectan un motivo RCGTG.  
 MUTREV-HRE: variantes que no afectan a ningún motivo RCGTG ni crean uno nuevo.

## N

NAMPT: Nicotinamide Phosphoribosyltransferase  
 NARF: Nuclear Prelamin A Recognition Factor  
 NCOA7: Nuclear Receptor Coactivator 7  
 NEW-HRE: variantes que crean un nuevo motivo RCGTG.  
 NGS: Next Generation Sequencing. Secuenciación masiva de segunda generación.  
 NHEJ: recombinación no homóloga o unión de extremos no homólogos  
 NTAD: Dominio de transactivación amino terminal.

## O

ODD: Dominio de degradación dependiente de oxígeno.  
 OSAS: Obstructive Sleep Apnea Syndrome

## P

P4HA2: Prolyl 4-Hydroxylase, Alpha Polypeptide II  
 PAM: motivo adyacente protoespaciador  
 PARN: Poly(A)-Specific Ribonuclease  
 PFKP: Phosphofructokinase, Platelet  
 PGK1: Fosfoglicerato kinasa 1.  
 PHD: Prolil Hidroxilasa  
 PKC: proteína quinasa C  
 PLC: fosfolipasa C

pO<sub>2</sub>: presión de oxígeno  
 PPME1: Protein Phosphatase Methylesterase 1  
 PPP1R3C: *Protein phosphatase 1 regulatory subunit 3C*.  
 PROL- : Prolactina

## Q

qPCR-RT: PCR cuantitativa a tiempo real

## R

RAP2B: Ras Family Small GTP Binding Protein  
 RCV: riesgo cardiovascular  
 REM: Rapid Eye Movement  
 RNMT: RNA (Guanine-7-) Methyltransferase  
 ROS: especies reactivas de oxígeno  
 RUNX1: Runt-related transcription factor 1  
 RUVBL2: RuvB-Like AAA ATPase 2

## S

SAHS: síndrome de apneas-hipopneas del sueño  
 SAOS: síndrome de apnea obstructiva del sueño  
 SCD: Stearoyl-CoA Desaturase  
 SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis.  
 SEC14L2: SEC14-Like 2 (*S. Cerevisiae*)  
 SEC61G: Sec61 Gamma Subunit  
 SHANK2: SH3 And Multiple Ankyrin Repeat Domains 2  
 SLC2A3: Solute Carrier Family 2 (Facilitated Glucose Transporter), Member 3  
 SLC35E1: Solute Carrier Family 35  
 SNP: Single Nucleotide Polymorphism  
 SUMF1: Sulfatase Modifying Factor 1

## T

TMEM18: Transmembrane Protein 18  
**V**  
 VEGFA: *Vascular endothelial growth factor A*.  
 VHL: *Von Hippel-Lindau*.

## W

WDR54: WD Repeat Domain 54  
 WDR89: WD Repeat Domain 89  
 WSB1: WD Repeat And SOCS Box Containing 1

## Z

**ZNF395:** *Zinc finger protein 395*.

4Su: 4-tiouridina

## **2 INTRODUCCIÓN**



## 2.1 Homeostasis del oxígeno

El oxígeno es una molécula imprescindible en el metabolismo aerobio de eucariotas, por ello estos organismos han desarrollado sistemas que permiten regular su homeostasis, manteniendo un fino equilibrio entre el aporte y consumo del mismo (Semenza 2004). Cuando la demanda de oxígeno excede al aporte se produce un desequilibrio que se denomina hipoxia.

Los organismos han desarrollado mecanismos de respuesta a la hipoxia a nivel sistémico y celular. A nivel sistémico, para contrarrestar los efectos agudos de la hipoxemia (disminución anormal de la presión parcial de oxígeno en sangre arterial por debajo de 80 mmHg), existen mecanismos que actúan rápidamente modulando la ventilación pulmonar y la perfusión, así como también la capacidad y la eficiencia del sistema cardiovascular para optimizar el suministro de oxígeno. Estos mecanismos están bajo el control de órganos y tejidos específicos entre los que se encuentran el cuerpo carotídeo en la circulación arterial, los cuerpos neuroepiteliales presentes en las vías aéreas y la respuesta directa de las células de músculo liso vascular en los pulmones (Gao et al. 2014).

Cuando la disminución del oxígeno se prolonga en el tiempo, las células animales ponen en funcionamiento un conjunto de respuestas celulares activando mecanismos que implican cambios en la expresión génica con el fin de reducir el consumo de oxígeno e incrementar su suministro. Esta respuesta, que se da en todos los tipos celulares estudiados, conlleva la regulación de la expresión de múltiples genes implicados en metabolismo, angiogénesis, eritropoyesis y ciclo celular, entre otras respuestas. En su conjunto, estas respuestas biológicas permiten la adaptación celular a la bajada de oxígeno y su puesta en marcha está mediada por el Factor Inducible por Hipoxia (HIF). Este factor de transcripción, que se expresa de forma ubicua en todos los tejidos, actúa a nivel celular como el regulador central de la respuesta a hipoxia. La identificación de HIF se realizó durante estudios encaminados a identificar la causa molecular de la inducción de eritropoyetina (*EPO*), una proteína que estimula proliferación y supervivencia de los progenitores de los eritrocitos de la sangre, en situaciones de hipoxemia tales como la anemia. Estudios iniciales demostraron que la cantidad de *EPO* en sangre aumentaba en condiciones de hipoxia y el análisis de los elementos en *cis* del gen de *EPO* llevó a la identificación del sitio de unión de HIF o elemento de respuesta a hipoxia (HRE), y la posterior caracterización de este factor transcripcional como regulador de *EPO* en condiciones de hipoxia (Semenza & Wang 1992).

La hipoxia puede ocurrir de manera continua o intermitente y a su vez, según su duración, puede ser aguda o crónica. Mientras que la hipoxia continua crónica ocurre tanto en contextos fisiológicos como patológicos, la hipoxia intermitente ocurre solamente en contextos patológicos. La hipoxia

como estímulo fisiológico ocurre en respuesta al crecimiento tisular durante el desarrollo o en algunos órganos como en el timo y en la médula ósea, donde la hipoxia es necesaria para el crecimiento y generación de nuevos timocitos (Hale et al. 2002) o para la progresión de la hematopoyesis (Cipolleschi et al. 1993), respectivamente. Por otro lado, la hipoxia está implicada en el desarrollo de procesos patológicos pudiendo actuar como potenciador o inhibidor de los mismos. En procesos tumorales, la expresión de HIF se ve aumentada en los núcleos hipóxicos formados durante el crecimiento tumoral (Semenza 2003). Además, HIF tiene un papel protector en enfermedades cardiovasculares que cursan con hipoxia basados en la reprogramación metabólica (Aragón et al. 2008). Así mismo se ha implicado a HIF en procesos de preconditionamiento hipóxico, este fenómeno se ha descrito en varios tejidos como corazón, cerebro y riñón, y supone un mecanismo protector en procesos de isquemia (Cai et al. 2007). Este factor de transcripción también está implicado en anemia, hemorragia y neumonía (Yuan et al. 2008). Por último, HIF tiene un rol central en enfermedades respiratorias que cursan con hipoxia como la Enfermedad Pulmonar Obstructiva Crónica (EPOC) y el Síndrome de Apnea Obstructiva del Sueño (SAOS).

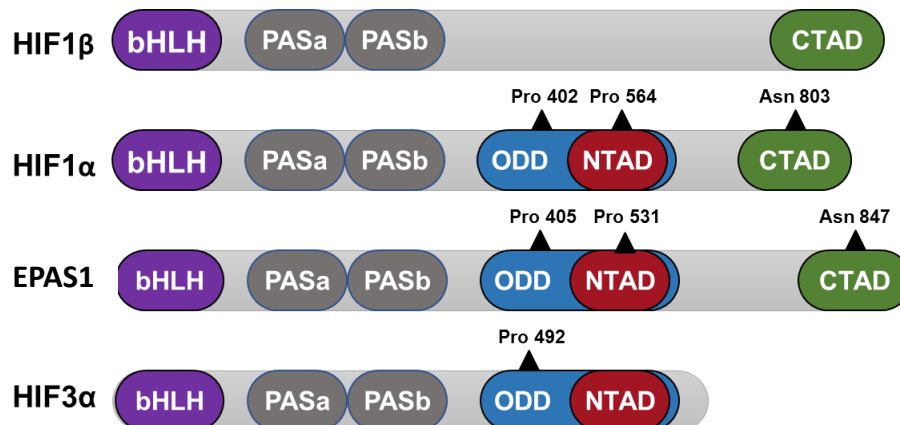
Puesto que HIF juega un importante papel en muchos procesos patológicos que cursan con hipoxia, el conocimiento global de la respuesta transcripcional mediada por este factor puede contribuir al desarrollo de nuevas estrategias terapéuticas dirigidas al tratamiento de estas patologías.

### 2.1.1 Estructura de HIF.

HIF es un factor de transcripción altamente conservado, identificado en múltiples especies de metazoos desde *Trichoplax adhaerens* (Loenarz et al. 2011) hasta *Homo sapiens*, lo que sugiere que su aparición fue un mecanismo adaptativo esencial en la evolución de los animales. La eliminación de HIF1A en ratones, resulta en letalidad embrionaria debido a fallos en el desarrollo placentario y el sistema circulatorio lo que pone de manifiesto el papel central de este factor en la formación de vasos durante el desarrollo (Iyer, Kotch, et al. 1998).

HIF es una proteína heterodimérica perteneciente a la familia de los factores de transcripción PER-ARNT-SIM (PAS) que es una subfamilia de la familia hélice-bucle-hélice básica (bHLH, del inglés basic helix-loop-helix). Está compuesto por una subunidad alfa, HIF $\alpha$ , cuya estabilidad depende de la tensión de oxígeno celular, y una subunidad beta, HIF $\beta$  (denominada ARNT, del inglés Aryl Hydrocarbon Translocator Factor), que se expresa de forma constitutiva (Hoffman et al. 1991). ARNT, además, puede dimerizar con otros factores de transcripción de la familia bHLH implicados en otras funciones como la regulación de los ciclos circadianos, crecimiento y señalización celular.(Yang et al. 2010).

Ambas subunidades tienen una elevada similitud estructural pues presentan en su extremo amino terminal un dominio bHLH que media la unión con ADN y en la región central contienen un dominio tipo PAS, que facilita la heterodimerización (Iyer, Leung, et al. 1998). En su extremo carboxilo terminal presentan el dominio de transactivación amino terminal (NTAD) y dominio de transactivación carboxilo terminal (CTAD) que les permiten interactuar con coactivadores como la ciclina dependiente de quinasa 8 (CDK8) y p300/CBP (Figura 1).



**Figura 1. Dominios estructurales de HIF.** Todas las subunidades de HIF comparten un dominio de unión al ADN, bHLH, y dos dominios de heterodimerización PAS. Las subunidades  $\alpha$  contienen un dominio ODD que contiene prolinas susceptibles a la hidroxilación por Prolil Hidroxilasas (PHDs). HIF1- $\alpha$  y EPAS-1 (también llamada HIF2- $\alpha$ ) presentan dos dominios de transactivación CTAD y NTAD, el dominio CTAD contiene una asparagina susceptible a la hidroxilación por Factor Inhibidor de HIF (FIH). HIF3- $\alpha$  y HIF1 $\beta$  contienen un solo dominio de transactivación

Existen tres isoformas de la subunidad HIF $\alpha$ : HIF1- $\alpha$ , EPAS-1 (también llamada HIF2- $\alpha$ ) y HIF3- $\alpha$ , que son codificadas por tres genes independientes. Todas ellas presentan un dominio de regulación dependiente de oxígeno (ODD), que contiene residuos de prolina (Pro402/Pro564 en HIF1- $\alpha$ , Pro405/Pro531 en EPAS-1 y Pro492 en HIF3 - $\alpha$ ) que determinan su estabilidad según la tensión de oxígeno (Figura 1). En el extremo carboxilo terminal presentan al menos un dominio de transactivación, NTAD y CTAD, uno en HIF3- $\alpha$  y dos en el caso de HIF1- $\alpha$  y EPAS-1, que incluyen una asparagina que puede ser hidroxilada (Asn 803 in HIF1- $\alpha$  and Asn 847 in EPAS-1) modificando la capacidad de HIF para reclutar P300/CBP, y por tanto determinando su actividad transcripcional (Jiang et al. 1997; Galbraith et al. 2013).

Como puede observarse en la figura 1 las subunidades HIF1- $\alpha$  y EPAS-1 presentan un mayor grado de semejanza estructural, lo que se refleja en un mayor solapamiento funcional. La subunidad HIF3- $\alpha$  es la peor caracterizada, ya que existen múltiples isoformas como resultado de procesamientos alternativos de su mensajero (Makino et al. 2001).

El heterodímero HIF $\alpha$ /HIF $\beta$  se une a la secuencia consenso RCGTG , (donde R indica purina, esto es: A/G) presente en sus genes diana (Wenger et al. 2005). Aunque HIF1-A y EPAS-1 unen el mismo motivo RCGTG, el conjunto de genes regulado por ambos factores no es completamente coincidente,

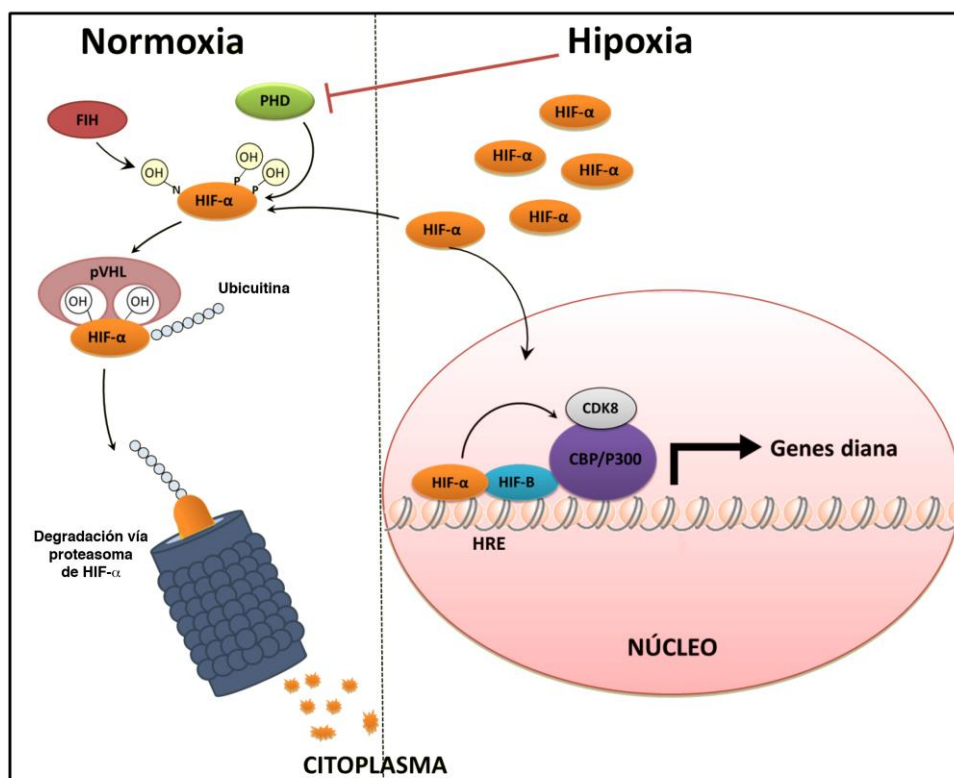
existiendo genes que solo se regulan por una de las dos isoformas. Esto puede deberse, entre otras razones, a la expresión diferencial de estas isoformas. Así, mientras que HIF1- $\alpha$  se expresa de forma ubicua EPAS-1 solo lo hace en unos pocos tejidos, siendo especialmente abundante en endotelio (de hecho se identificó en este tipo celular y por ello recibe el nombre del inglés *Endothelial PAS domain 1*), hígado, riñón, pulmón y cerebro (Tian et al. 1997). Por otro lado, ambas isoformas interaccionan con diferentes factores de transcripción como proteína activadora 1 (AP-1), y factor activador de transcripción 1 (ATF1) (Firth et al. 1995). Lo que sin duda contribuye al diferente patrón transcripcional regulado por estos factores de transcripción.

### 2.1.2 Regulación de la actividad transcripcional de HIF.

Se han propuesto numerosos mecanismos moleculares para explicar la activación de HIF en respuesta a la variación de la presión de oxígeno ( $pO_2$ ) (Semenza 1999). El modelo establecido propone que las prolil hidroxilasas (PHDs) o EGL-*nine* homólogos (EGLNs) son los sensores celulares de oxígeno que regulan la actividad transcripcional de HIF a nivel del control de la estabilidad de las subunidades alfa. Las EGLNs, son enzimas que utilizan oxígeno molecular para catalizar la hidroxilación de residuos de prolina situados en el dominio ODD de HIF. La hidroxilación en al menos uno de estos residuos determina la estabilidad de HIF- $\alpha$ , ya que permite su reconocimiento por parte de la proteína Von Hippel-Lindau (VHL) que forma parte de un complejo E3-ubiquitina ligasa que marca a HIF- $\alpha$  para su posterior degradación vía proteasoma (Maxwell et al. 1999) (Figura 2). Las EGLNs pertenecen a una familia de dioxigenasas que utilizan oxígeno y 2-Oxoglutarato como co-sustratos, y hierro y ascorbato como cofactores para catalizar la reacción de hidroxilación, pero su característica más importante es que poseen una  $K_m$  para el oxígeno muy elevada, 100-250  $\mu M$  que está en el rango de la concentración de oxígeno en disolución a presión atmosférica (280  $\mu M$ ), de modo que pequeñas variaciones de la  $pO_2$  en la célula provocan una disminución drástica de su actividad enzimática (Hirsilä et al. 2003). Como consecuencia de ello, en hipoxia, HIF no es hidroxilado, escapa a la degradación por VHL y se trasloca al núcleo (Ehrismann et al. 2007). Una vez en el núcleo dimeriza con ARNT, cuya expresión es constitutiva, y promueve la regulación transcripcional mediante la unión al ADN de sus genes diana que están implicados en la adaptación a la bajada de la  $pO_2$  (Figura 2).

Una vez que se ha formado el heterodímero HIF $\alpha$ /HIF $\beta$  en el núcleo, para formar un complejo transcripcionalmente activo, se requiere el reclutamiento de diversos coactivadores que tienen como función estabilizar el complejo de iniciación de la transcripción y promover la acetilación de histonas que favorece la apertura de la cromatina. El principal complejo coactivador es p300/CBP, que interacciona con el dominio CTAD de HIF- $\alpha$  ayudando al posterior reclutamiento del resto de coactivadores (Freedman et al. 2003).



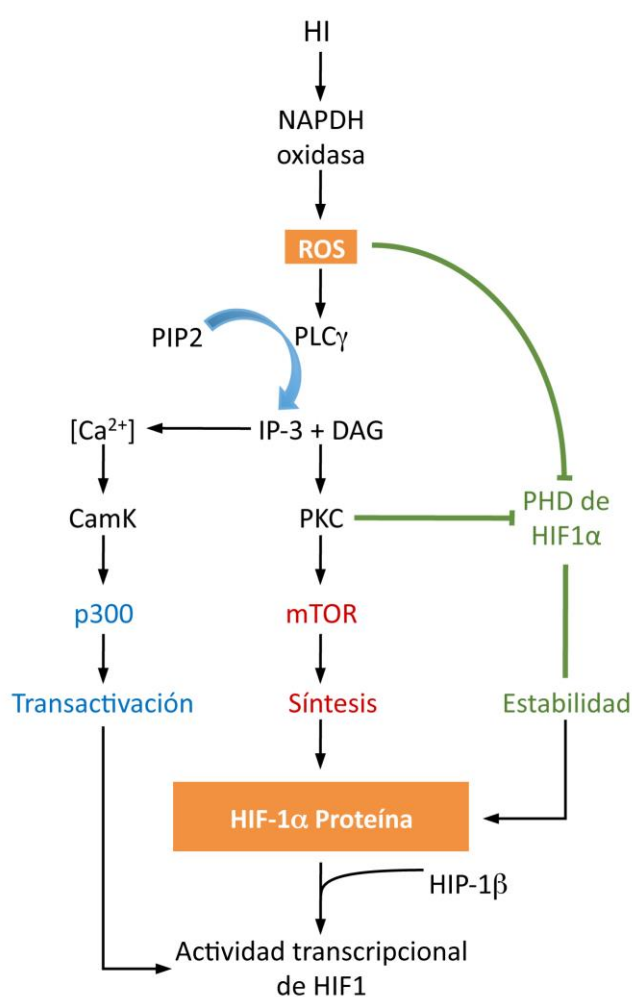


**Figura 2. Regulación de la actividad transcripcional de HIF.** En normoxia las PHDs hidroxilan las subunidades HIF- $\alpha$  en los residuos de prolina de los dominios ODD, permitiendo el reconocimiento por la proteína VHL que forma parte del complejo E3 ubiquitina ligasa que lo marca para su degradación vía proteasoma. Además, FIH hidroxila la asparagina presente en el dominio CTAD, evitando el reclutamiento de CBP/P300. Cuando ocurre un desequilibrio entre el aporte y la demanda de oxígeno (hipoxia), las prolil hidroxilasas (PHDs y FIH) ven comprometida su actividad, HIF- $\alpha$  se estabiliza y se trasloca al núcleo donde dimeriza con ARNT. El factor de transcripción activo se une a los elementos de respuesta a hipoxia (HRE, cuya secuencia consenso es RCGTG) y recluta coactivadores como CDK8 y CBP/P300, induciendo la expresión de sus genes diana (Adaptado de (Maes et al. 2012)).

La regulación de HIF más estudiada corresponde a la hipoxia continua o crónica, explicada previamente, que se da en el caso de los pacientes con EPOC. Los períodos de hipoxia breve (15-30 segundos) y reoxigenación (5min) dan lugar a hipoxia intermitente, tal como ocurre en SAOS, donde la actividad transcripcional de HIF se induce por mecanismos poco conocidos y diferentes a los explicados previamente (Figura 3). Así, la hipoxia intermitente activa la enzima NADPH oxidasa produciendo especies reactivas de oxígeno (ROS) que inducen la actividad de la fosfolipasa C (PLC), dando lugar a la generación de inositol 3 fosfato (IP3) y diacilglicerol (DAG). La unión de IP3 a su receptor lleva a la movilización de  $\text{Ca}^{+2}$  intracelular y junto con DAG activan la proteína quinasa C (PKC). PKC estimula la acumulación de HIF1- $\alpha$ , por un lado aumentando su síntesis a través de mTOR (de sus siglas en inglés, mammalian Target of Rapamycin) y por otro lado inhibiendo su degradación dependiente de PHD. Además, el aumento de  $\text{Ca}^{+2}$  intracelular activa la calcio-calmodulina quinasa (CamK) aumentando la actividad del coactivador de HIF, p300. Por tanto, la

hipoxia intermitente desencadena vías de señalización que conducen al aumento de la actividad transcripcional de HIF1- $\alpha$  (Yuan et al. 2008).

En contraste con la hipoxia crónica, en la que HIF1- $\alpha$  se degrada rápidamente ( $t_{1/2} < 5$  minutos) tras la reoxigenación, los niveles de HIF1- $\alpha$  se mantienen elevados después de la hipoxia intermitente debido a la activación mantenida de mTOR, lo cual resulta significativo en el contexto de SAOS donde la respuesta patológica respiratoria y cardiovascular permanecen durante horas después del cese de la hipoxia intermitente (Yuan et al. 2008).

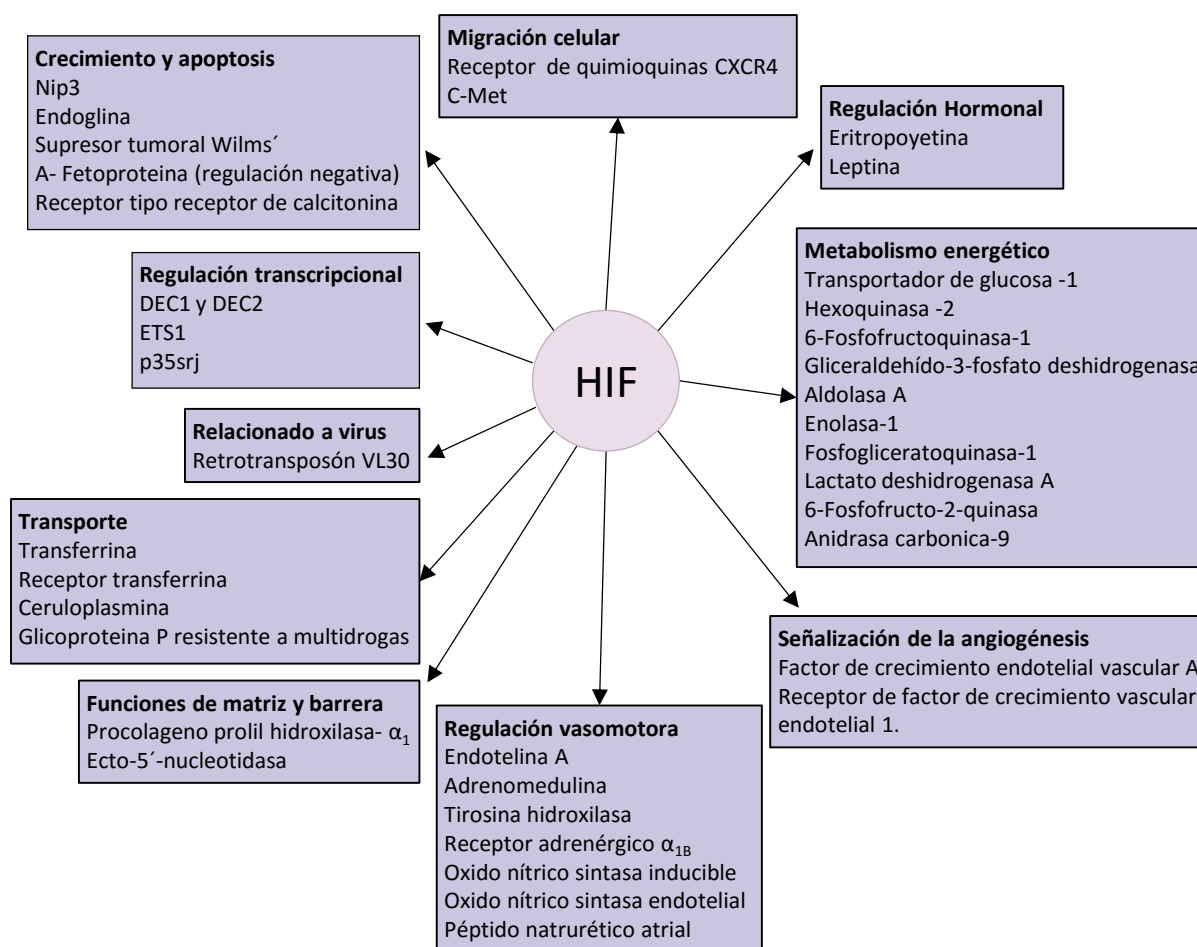


**Figura 3. Regulación de la actividad transcripcional de HIF en hipoxia intermitente.** Se indican las vías de señalización mediante las que la hipoxia intermitente induce HIF1- $\alpha$ . Esta activación se inicia con la generación de ROS por NADPH oxidasa, que por distintas vías de señalización culmina en el aumento de la actividad transcripcional de HIF1- $\alpha$ , ya sea aumentando la síntesis (en rojo) y la transactivación (azul) o evitando su degradación y por tanto aumentando la estabilidad proteica (verde) (Adaptado de (Yuan et al. 2008)).

### 2.1.3 Regulación de la expresión génica por HIF.

Independientemente de cómo se ha inducido la actividad transcripcional de HIF1- $\alpha$ , ya sea por hipoxia crónica o hipoxia intermitente, las células responden a cambios en la disponibilidad de oxígeno que permiten la adaptación a hipoxia. Esta respuesta conlleva la expresión de varios genes implicados en múltiples funciones tales como: el metabolismo energético, angiogénesis, regulación transcripcional, migración celular, ciclo celular y apoptosis, entre otros procesos indicados en la [Figura 4](#). Globalmente se pueden diferenciar dos tipos de respuestas, una orientada a aumentar el

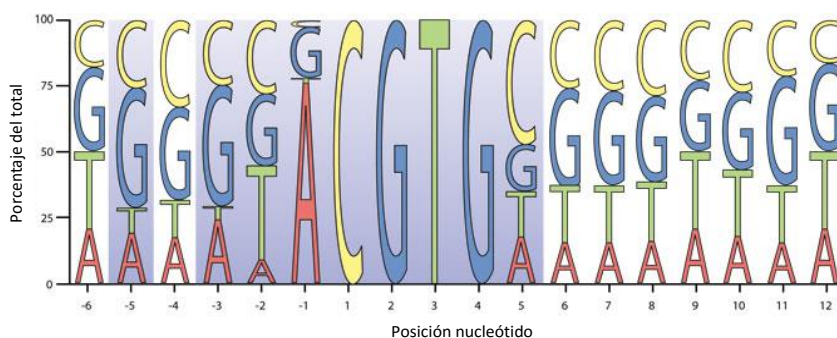
aporte de oxígeno a regiones hipóxicas y otra que busca reducir el consumo de oxígeno reprogramando el metabolismo celular (Jiang et al. 1997).



**Figura 4. Rutas biológicas inducidas por HIF.** Se han descrito más de 100 genes diana de HIF implicados en diversas funciones biológicas que permiten la adaptación celular a la bajada de oxígeno. (Adaptado de (Schofield & Ratcliffe 2004)).

El estudio de la regulación de genes específicos mediado por HIF, llevó a la identificación de una secuencia consenso cuyo núcleo central es el motivo RCGTG (Wenger et al. 2005) y se ha puesto de manifiesto la importancia de las posiciones adyacentes al núcleo central del HRE a la hora de hacer distinciones entre HREs funcionales y no funcionales. Posteriormente, se ha confirmado que HIF se une al ADN a través de la interacción con este motivo mediante el análisis global del patrón de transcripción inducido por hipoxia con técnicas de alta resolución (Schodel et al. 2011). En la [figura 5](#) se ve reflejada la frecuencia de aparición de cada base en cada posición en torno al RCGTG, así vemos que existen bases que aparecen con mayor frecuencia que la esperada por azar, determinando que la secuencia HRE es la comprendida en las 18 pares de bases que contienen al núcleo central RCGTG. Se ha visto que la presencia de este núcleo central o HRE es necesaria pero no suficiente para la

transcripción de los genes diana de HIF en respuesta a hipoxia. Lo cual se evidencia en la existencia de una serie de regiones en torno al HRE necesarias para su actividad transcripcional a las que se unen otros factores de transcripción que colaboran con HIF, y que no necesariamente se inducen por hipoxia (Ebert & Bunn 1998). Estos factores pueden potenciar la actividad transcripcional de HIF y determinar regulación específica de tejido.



**Figura 5. Secuencia consenso del elemento de respuesta a HIF (HRE).** En el logo quedan reflejadas las posiciones en torno al núcleo central del HRE, RCGTG, y la frecuencia de aparición de cada base nucleotídica en cada posición (Adaptado de (Wenger et al. 2005)).

En relación a la importancia de la composición del núcleo central RCGTG, nuestro grupo demostró que la presencia de alteraciones en la secuencia del HRE puede impedir la regulación de los genes diana. En concreto, se demostró que el polimorfismo de un único nucleótido (SNP), rs17004038, localizaba en un HRE funcional en el locus del factor inhibidor de macrófagos (MIF) y que la presencia de este polimorfismo impedía la inducción de MIF por hipoxia (Ortiz-Barahona et al. 2010). Este resultado, implica que diferentes individuos van a responder a la hipoxia de forma desigual en función de las variantes genéticas que porten en sus genomas.

## 2.2 Enfermedades respiratorias

Las enfermedades que afectan el sistema respiratorio tienen un efecto sistémico significativo puesto que no solo afectan a nivel pulmonar sino también a nivel extra respiratorio. Muchas de ellas convergen en un intercambio gaseoso ineficaz y, por tanto, una disponibilidad de oxígeno ineficiente que dificulta el normal funcionamiento de los tejidos que dependen de un metabolismo aerobio.

Las enfermedades respiratorias crónicas afectan las vías respiratorias y otras estructuras del pulmón. Algunas de las más frecuentes son: el asma, la EPOC, las alergias respiratorias, las enfermedades pulmonares de origen laboral, el SAOS y la hipertensión pulmonar. En particular, para este trabajo nos centramos en el estudio de la EPOC y el SAOS que se caracterizan por el desarrollo de una situación hipóxica. Mientras que los pacientes con EPOC sufren una hipoxia sostenida crónica,

el SAOS se caracteriza por el desarrollo de una hipoxia intermitente crónica (Prabhakar & Semenza 2012).

### **2.2.1 Enfermedad pulmonar obstructiva crónica (EPOC)**

La EPOC se caracteriza esencialmente por una limitación crónica al flujo aéreo asociada a una respuesta inflamatoria exagerada de las vías aéreas y del parénquima pulmonar frente a partículas o gases nocivos (Calle et al. 2012). Los síntomas de EPOC incluyen disnea, tos y expectoración crónica, y se asocia a dos enfermedades principales: el enfisema pulmonar y la bronquitis crónica. El enfisema pulmonar consiste en el agrandamiento permanente de los bronquiolos terminales con destrucción de la pared alveolar que provoca el colapso de las vías respiratorias, pudiendo o no presentarse fibrosis. La bronquitis crónica consiste en la inflamación de los bronquios, que hace que se reduzca el flujo de aire de entrada y salida de los pulmones, y un aumento de la secreción de mucus que obstruye las vías respiratorias.

Los pacientes con EPOC severa pueden presentar hipoxemia e hipercapnia (aumento de la presión parcial de dióxido de carbono en sangre arterial por encima de 46 mmHg), con un empeoramiento del intercambio de gases durante el sueño, pues hipoventilan, provocando desaturaciones nocturnas y retención de dióxido de carbono (Halbert et al. 2006).

El riesgo de padecer EPOC se relaciona con la carga total de partículas inhaladas a las que un individuo se expone durante su vida, siendo el principal factor de riesgo el tabaquismo. Otros factores de riesgo de importancia son la polución de ambientes interiores, contaminación atmosférica y exposición ocupacional (Soriano & Lamprecht 2012; Mannino & Buist 2007). Por otra parte, existen factores genéticos que predisponen a esta enfermedad. El factor de riesgo genético más conocido es la deficiencia hereditaria de  $\alpha$ 1-antitripsina ( $\alpha$ 1-ATD) (odds ratio [OR] =4,39; Intervalo de confianza 95% [IC95%]= 4,20- 5,15) que está presente en el 3% de los pacientes con EPOC (Ganrot et al. 1967; Stoller & Aboussouan 2005; Mcgrady et al. 2015). Por otro lado, se ha identificado un SNP rs8034191 en el receptor nicotínico colinérgico alfa 5/3 asociado a EPOC (OR= 1,404; pvalor= 0,0001) y cuyo riesgo en la población es de 12,2%, representando el locus de mayor contribución a la EPOC en la población general (Pillai et al. 2009).

El diagnóstico de EPOC se realiza mediante una espirometría forzada, que mide la capacidad funcional pulmonar, para evaluar la gravedad de la enfermedad. Así, se diagnostica con EPOC a los individuos que presentan un volumen espiratorio forzado en el primer segundo (FEV1) post-broncodilatador menor al 80 % del valor teórico y una relación entre FEV1 y la capacidad vital forzada (FVC), menor a 0,70 (Halbert et al. 2006).

Existen diferentes variantes de la enfermedad o fenotipos que incluyen: enfisema, bronquitis crónica, agudizador con bronquitis crónica, agudizador con enfisema y mixto (EPOC-asma). Las agudizaciones o exacerbaciones son episodios de inestabilidad clínica que precisan al menos de tratamiento con corticoesteroides sistémicos y/o antibióticos (Calle et al. 2012), y se define como agudizador a todo paciente con EPOC que presente 2 o más agudizaciones al año. El fenotipo mixto se refiere a pacientes que presentan características de más de una enfermedad obstructiva de las vías, es decir tiene un síndrome de solapamiento (Miravittles et al. 2013). Los procesos o enfermedades más representativas y frecuentes son la obstrucción crónica al flujo aéreo y el asma. Además de estos fenotipos relativos a la EPOC *per se*, los pacientes pueden presentar una o más enfermedades que coexisten independientemente de la enfermedad primaria y se denominan comorbilidades. Las comorbilidades que se asocian con mayor frecuencia a EPOC son: hipertensión arterial, diabetes mellitus, dislipemia, enfermedades psiquiátricas (depresión y ansiedad), trastornos gastrointestinales, artritis, osteoporosis, infecciones, cáncer y enfermedades cardiovasculares (de Miguel Díez et al. 2010). Las comorbilidades en la EPOC no sólo contribuyen a aumentar la repercusión social y el coste anual de la enfermedad, sino que también constituyen un factor pronóstico de mortalidad en los pacientes (de Miguel Díez et al. 2010).

La EPOC es una enfermedad muy heterogénea y de difícil clasificación. Por ello la estratificación de los pacientes de EPOC según las escalas de gravedad o pronóstico se ha realizado según varios criterios:

- i. GOLD, de *Global Initiative for Chronic Obstructive Lung Disease*: se categoriza la EPOC según la gravedad de la limitación al flujo aéreo (a partir de la FEV1 post-broncodilatador), en pacientes con cociente FEV1/FVC post-broncodilatador  $< 0,70$  se clasifica en :
  - GOLD 1: Leve, FEV1  $> 80$  % del valor de referencia teórico.
  - GOLD 2: Moderada,  $50\% < \text{FEV1} < 80\%$  del valor de referencia teórico.
  - GOLD 3: Grave,  $30\% < \text{FEV1} < 50\%$  del valor de referencia teórico.
  - GOLD 4: Muy grave, FEV1  $< 30$  % del valor de referencia teórico
- ii. BODE/BODEx, de *Body-mass index, airflow Obstruction, Dyspnea, and Exercise capacity index*: clasificación de la enfermedad según la gravedad, para ello se combinan varios parámetros como el propio FEV1, la distancia que el sujeto puede recorrer en 6 minutos, el valor de la escala de disnea (MMRC, Modified Medical Research Council dyspnea Scale), y el índice de masa corporal.
- iii. COTE, de *COPD specific CO-morbidity TEst*. Es un índice que a diferencia de los anteriores tiene en cuenta las comorbilidades que se asocian a una mayor mortalidad, como es el caso de cáncer (en particular de pulmón, esófago, páncreas y mama), ansiedad, cirrosis hepática, fibrilación auricular, diabetes, fibrosis pulmonar, insuficiencia cardíaca, úlcera gastroduodenal y

enfermedad coronaria. Otras comorbilidades de gran frecuencia en pacientes con EPOC como hipertensión e hipercolesterolemia no estarían asociadas a mayor mortalidad.

La EPOC constituye un problema socio sanitario de primera magnitud debido a su elevada prevalencia y morbimortalidad. Se estima que hay 210 millones de personas afectadas por EPOC en el mundo y que causa la muerte de 2,9 millones de personas anualmente (Soriano & Lamprecht 2012), en particular, en España la EPOC afecta al 10,2% de la población adulta (Miravitlles et al. 2009).

Una de las dificultades en el tratamiento del EPOC es la heterogeneidad que se observa en el desarrollo de los fenotipos de los pacientes. Este hecho se pone de manifiesto en los pacientes fumadores que desarrollan la enfermedad presentando manifestaciones clínicas muy variables con predominio de eritrocitosis, hipertensión pulmonar, complicaciones cardiovasculares o agudizaciones frecuentes (Mannino et al. 2006). La explicación a dicha variabilidad fenotípica puede ser debida en parte a la existencia de variantes genéticas que predisponen a complicaciones específicas de la enfermedad. Por ejemplo, las variantes identificadas en  $\alpha 1$ -ATD disminuyen su función inhibidora provocando que el tejido conectivo fibroso pulmonar sea más sensible al daño por la elastasa que actúa en las paredes alveolares alterando su estructura (Berndt et al. 2012). Además se han descrito dos polimorfismos 3A4A y 4A4A en el gen que codifica para endotelina 1 asociados a un riesgo mayor de padecer EPOC (OR = 1,427; IC95% = 1,089–1,871 y OR = 2,622; 95%IC = 0,842–8,165, respectivamente, versus el genotipo de referencia:3A3A) (Sampsonas et al. 2010). Otros polimorfismos identificados en genes que codifican para enzimas que participan en la neutralización de radicales libres y detoxificantes del estado redox provocado en EPOC podrían tener efectos en la progresión de la enfermedad (Lakhdar et al. 2011). Además se ha descrito un SNP rs2276109 en la metaloproteinasa *MMP12* que se asocia con un menor riesgo de padecer EPOC en fumadores (OR= 0,63; IC95% = 0,45–0,88) (Hunninghake et al. 2009). Posteriormente, en un estudio del genoma completo se han analizado SNPs en relación a FEV1 y el cociente FEV1/ FVC, identificando 16 locus que no habían sido relacionados con el EPOC previamente y están implicados en rutas biológicas que regulan la función pulmonar (Artigas et al. 2011).

Estos estudios ponen de manifiesto que la base genética de esta enfermedad es de carácter multifactorial y por tanto es la interacción de varios genes con efectos pequeños la responsable de los fenotipos observados en la clínica (Chen et al. 2013; Regan et al. 2011). Por ello, es de esperar que existan muchos factores genéticos aún no identificados que afecten la aparición y/o desarrollo de la enfermedad (Silverman et al. 2011).



### 2.2.2 Síndrome de apnea obstructiva del sueño (SAOS)

El SAOS, también nombrada Síndrome de Apneas-Hipoapneas del Sueño (SAHS), se caracteriza por episodios recurrentes de colapso total o parcial de las vías aéreas superiores durante el sueño, originando episodios de hipoxia intermitente y fragmentación del sueño. La apnea es una interrupción completa y temporal de la respiración de entre 10-40 segundos de duración, mientras que hipoapnea es la reducción del flujo de aire hacia los pulmones; en ambos casos se produce una disminución de los niveles de oxígeno y aumento de dióxido de carbono ( $\text{CO}_2$ ) en sangre, conjuntamente con microdespertares subconscientes que permiten reanudar la respiración hasta que se produce el siguiente episodio. Estas apneas y/o hipoapneas pueden provocarse por la relajación de los tejidos blandos de la parte posterior de la garganta o por la estrechez de ésta, lo cual bloquea el paso del aire en las vías aéreas superiores (apnea obstructiva) o puede ocurrir por una disfunción de los centros respiratorios en los que el sistema nervioso central deja de enviar el estímulo respiratorio (apnea central) (Nanduri et al. 2015). La frecuencia de las apneas puede variar, alcanzando hasta 60 episodios por hora en pacientes severamente afectados, en los que la saturación de  $\text{O}_2$  de la hemoglobina en sangre puede reducirse hasta el 50% (Prabhakar & Semenza 2012).

Además de la fragmentación del sueño y la hipoxemia nocturna, estos pacientes presentan: ronquido estruendoso, sueño no reparador con despertares frecuentes, obesidad, cefalea matutina e hipertensión arterial; muchos de estos síntomas están asociados a la activación del sistema simpático producida por las apneas (Lloberes et al. 2011). Las secuelas a largo plazo del SAOS incluyen el incremento en la morbilidad cardiovascular, en la mortalidad, empeoramiento en las funciones neuropsicológicas y excesiva somnolencia diurna (Findley et al. 1986; Prabhakar & Semenza 2012).

El mayor factor de riesgo en SAOS es la obesidad ( $\text{IMC} > 35$ ), seguido del sexo siendo más prevalente en hombres que en mujeres (3:1) y la edad, la población adulta de entre 40-65 años es la más afectada. El alcohol, tabaco, sedantes, hipnóticos y barbitúricos pueden influenciar la aparición o agravamiento de SAOS, así como también es frecuente en individuos que presentan anomalías faciales (Lloberes et al. 2011).

En cuanto al diagnóstico de SAOS, los episodios de apnea se miden con el Índice de Apneas-Hipoapneas (IAH) o Índice de Alteración Respiratoria (IAR), que consiste en sumar el número de apneas, hipoapneas y Esfuerzos Respiratorios Asociados a Microdespertares (ERAM) durante una hora de sueño. Se diagnostica SAOS cuando el índice IAR es mayor a 5 y se encuentra asociado Excesiva Somnolencia Diurna (ESD) no explicada por otras causas, o cuando el índice  $\text{IAR} > 5$  está



asociado a dos o más de los siguientes síntomas: asfixias repetidas durante el sueño, despertares recurrentes durante el sueño, percepción del sueño como no reparador, cansancio y/o fatiga durante el día y dificultades de concentración (Lloberes et al. 2011).

Los pacientes con SAOS se clasifican según la gravedad, teniendo en cuenta la frecuencia de las apneas o hipoapneas durante una hora (McNicholas 2008):

- Leve, cuando la frecuencia es de entre 10 y 20 veces por hora.
- Moderado, cuando la frecuencia es de entre 20 y 30 veces por hora.
- Severo, si la frecuencia supera las 30 veces por hora.

Las comorbilidades más frecuentemente asociadas a esta enfermedad son hipertensión arterial (HTA), síndrome metabólico, insuficiencia cardíaca, hipertensión pulmonar, tromboembolismo pulmonar, cáncer, depresión y ansiedad, dislipidemia, diabetes mellitus y cardiopatía isquémica. En concreto, además de originar una mayor somnolencia diurna, se ha demostrado que el SAOS se asocia con un importante incremento de la mortalidad cardiovascular (Marin et al. 2005; Shahar et al. 2001; Peppard et al. 2000). Esta comorbilidad cardiovascular estaría favorecida por el estado pro-inflamatorio sistémico que se genera como consecuencia de los episodios nocturnos de oxigenación-reoxigenación que inducen las apneas-hipoapneas (Gozal & Kheirandish-Gozal 2008).

La prevalencia de SAOS en la población adulta a nivel global es de un 2-4% (Young et al. 1993), mientras que en España afecta a un 2,2% de los hombres y un 0,8% de las mujeres (Marin et al. 1997).

EL SAOS, al igual que la EPOC es una enfermedad multifactorial, lo cual se describe en estudios epidemiológicos donde se observa que la varianza del índice IAH puede ser explicada por causas genéticas (Redline et al. 1995) y en segundo lugar en la identificación de asociaciones entre factores genéticos y comorbilidades como la obesidad, la morfología cráneo facial y el control de la ventilación y el desarrollo del síndrome (Zhao et al. 2013) . Además, existe una relación recíproca entre SAOS y diabetes mellitus (Cerón et al. 2015). Por un lado, SAOS se asocia con la resistencia a insulina e intolerancia a la glucosa y por otra parte varios estudios han demostrado una elevada prevalencia de SAOS en pacientes con diabetes tipo 2. Otro gen vinculado a SAOS es *VEGF*, donde se describió que el alelo -460 C (rs833061) (OR = 1,95, IC95% = 1,38-2,76) y el alelo +936 T (rs3025039) (OR = 1,48, IC95% = 1,02-2,15) se asocian a una mayor susceptibilidad de padecer SAOS (Cao et al. 2014).

En otro trabajo en el que se estudió el receptor de serotonina *5-HTR 2A*, involucrado en la neurotransmisión de distintos procesos biológicos entre ellos el sueño, se ha identificado que los individuos que poseen el genotipo A/A en -1438G tienen un riesgo aumentado de SAOS (OR = 3,69, IC95% = 1,92-7,08) comparado con aquellos individuos que poseen el genotipo G/G (Zhao et al.

2013). Recientemente se han identificado variantes genéticas en el promotor de la interleucina 10 que se asocian a mayores riesgos de padecer SAOS concretamente, los SNPs rs1800896, rs1800871 y rs1800872 (OR= 1,258, IC95% = 1,187–1,329; OR=1,247, IC95% = 1,177–1,316 y OR=1,159 IC95% = 1,101–1,216, respectivamente) (Özdaş et al. 2015). La presencia de estos polimorfismos puede explicarse porque la interleucina 10, además de su papel en inflamación, está implicada en la regulación de los patrones normales del sueño, dos características que se ven afectadas en los pacientes con SAOS.

### 2.2.3 Papel de HIF en EPOC y SAOS.

La EPOC y el SAOS se caracterizan por el desarrollo de una situación hipóxica, que es continua en la EPOC e intermitente en el SAOS, pero que en ambos casos conlleva la inducción de un programa génico específico bajo el control de HIF (Prabhakar & Semenza 2012).

El estudio de modelos animales de EPOC sugiere que HIF es necesario para la progresión de la enfermedad. Ratones heterocigotos para EPAS-1 expuestos a hipoxia crónica están protegidos contra el desarrollo de hipertensión pulmonar e hipertrofia del ventrículo derecho, dos comorbilidades observadas en pacientes de EPOC, respecto de los ratones salvajes (Brusselmans et al. 2003). Por otro lado, se pudo demostrar que la expresión de EPAS-1 está elevada en neutrófilos infiltrados en el pulmón de pacientes con EPOC moderado y severo respecto de individuos no fumadores (Thompson & Elks 2014), demostrando que HIF se activa. De hecho, la inducción de HIF en pacientes con EPOC, podría explicar la elevación del hematocrito y la inflamación sistémica (Kent et al. 2011). De acuerdo con esta posibilidad, un estudio reciente en el que se ha comparado la expresión génica y el estado de metilación de la cromatina de muestras de pulmón obtenidas de pacientes y de individuos control, se ha identificado a EPAS-1 como un regulador central de la expresión génica en EPOC (Yoo et al. 2015).

Por otro lado, en relación al papel de HIF en SAOS, se ha descrito que la hipoxia intermitente, promueve la acumulación de HIF1-alpha en células adrenales de ratón (Yuan et al. 2008), contribuyendo a un aumento del estrés oxidativo mediante la inducción de enzimas pro-oxidantes (Prabhakar & Semenza 2012). Además, mediante el uso de ratones heterocigotos para HIF1-alpha sometidos a hipoxia intermitente, se demostró que la respuesta cardiorespiratoria mediada por el cuerpo carotídeo requiere la actividad de HIF (Peng et al. 2006).

En conjunto, todos estos trabajos demuestran la activación de HIF en la EPOC y el SAOS, y sugieren que puede jugar un papel central en el desarrollo de algunas complicaciones de estas enfermedades.

### 2.3 Papel de las variantes genéticas en regiones reguladoras en la variabilidad fenotípica interindividuo.

En los últimos años, gracias a los avances en las técnicas de alto rendimiento (microarrays genómicos y secuenciación masiva) y la disminución de su coste, se ha realizado un gran número de estudios de asociación de genoma completo (en inglés, *GWAS Genome-wide association study*). En este tipo de estudios se identifican todas las variantes a lo largo del genoma en una población concreta y se estudia si alguna de ellas correlaciona de forma significativa con un fenotipo de interés.

De entre los múltiples tipos de variaciones genéticas que pueden encontrarse en el genoma humano los SNPs, que son variaciones de secuencias cuya frecuencia es mayor al 1% en la población general, constituyen el tipo más común de polimorfismo. Su frecuencia en el genoma humano está estimada en 1 SNP cada 100-300 pares de bases lo que supone aproximadamente 15 millones de SNPs (The 1000 Genomes Project Consortium 2010).

Aunque la mayoría de los SNPs no son causantes de enfermedades ni están asociados a patologías, ha sido posible clasificarlos según estén asociados a un fenotipo o tengan efectos funcionales. Distinguiendo entre:

- i) SNPs indicativos, asociados a una respuesta diferencial a algún fármaco o al riesgo de padecer cierta enfermedad sin explicación causal. No se localizan en el cuerpo de un gen ni afectan la función proteica.
- ii) SNPs en regiones codificantes, provocan variaciones en la secuencia del ADN en regiones codificantes, pudiendo o no tener efectos según se trate de un SNP silencioso o según se trate de un SNP que provoque un cambio en la identidad de los aminoácidos.
- iii) SNPs en regiones no codificantes asociados a secuencias reguladoras, modifican la expresión génica ya sea temporalmente, en su localización o en el nivel de la expresión.

La posibilidad de estudiar variantes a lo largo de todo el genoma, ha aumentado la atención sobre variaciones genéticas en regiones no codificantes como las responsables de cambios en la regulación génica en múltiples enfermedades (Mathelier et al. 2015). En un meta-análisis en el que se estudiaron aproximadamente 1.200 SNPs para los que se había encontrado asociación con alguna enfermedad humana, se observó que el 40% de ellos se encontraron fuera de exones poniendo en evidencia la importancia de los SNPs en regiones no codificantes, así como su implicación en patologías (Visel et al. 2009).

Muchos de los SNPs en regiones no codificantes asociados a patologías, se localizan en sitios de unión de factores de transcripción, ya sea en las proximidades o a varias kilobases del gen que regulan. Un ejemplo clásico de regulación a corta distancia es el caso de la variante de un único

nucleótido (cambio de una T a una A) en el promotor del Factor IX endopeptidasa (*F9*), un gen involucrado en la coagulación sanguínea. Este SNP se encuentra en la posición más conservada del sitio de unión del factor de transcripción HNF4A y su presencia provoca una menor actividad del promotor de *F9* causando la hemofilia B de Leyden (Reijnen et al. 1993). Un ejemplo de variante que localiza a varias kilobases del gen regulado es el SNP rs4988235 que provoca un cambio de secuencia de C/T localizado a 14 kilobases (-13910) del locus de la enzima lactasa (lactosa-phlorizin-hidrolasa) que resulta en una menor actividad de esta enzima en la células intestinales provocando intolerancia a la lactosa (Enattah et al. 2002). Las variantes en regiones no codificantes pueden afectar a la expresión de genes localizados a grandes distancias, lo que dificulta enormemente el análisis de su impacto funcional. Por ejemplo, en los intrones 1 y 2 del gen FTO existen varios SNPs que constituyen la asociación genética más fuerte conocida con obesidad en humanos. Pese a su localización en el locus de FTO, recientemente se ha descrito que estos polimorfismos afectan a la expresión de IRX3 localizado a más de 500kb de los SNPs (Smemo et al. 2014).

Todos estos trabajos describen ejemplos donde la presencia del SNP resulta en una disminución en la transcripción, pero también hay casos en los que un polimorfismo puede crear un sitio de unión para un factor de transcripción aumentando la expresión del gen en que se encuentra. Por ejemplo, el polimorfismo en el promotor de la metaloproteasa I (*MMPI*), donde la inserción de una G crea un sitio de unión para el factor de transcripción Ets (GGA) y resulta en un aumento en la transcripción de MMPI en fibroblastos y células de melanoma (Rutter et al. 1998). Puesto que MMPI es central en la degradación del estroma y la consecuente invasión tumoral, la presencia de este polimorfismo podría repercutir en la progresión del cáncer. Por tanto, la presencia de un polimorfismo en regiones reguladoras puede tener efectos en la regulación génica, ya sea disminuyendo o aumentando la actividad del promotor, que pueden afectar a la progresión de diferentes patologías.

En relación a EPOC, si bien se han realizado estudios en los que se han identificado numerosas variantes asociadas a la enfermedad (Berndt et al. 2012; Chen et al. 2013), existen pocos estudios en los que además de la asociación se estudie el efecto funcional del polimorfismo. En particular, para este trabajo nos interesan aquellos SNPs que afectan sitios de unión de HIF alterando características o síntomas de los pacientes con EPOC y SAOS pero hasta la fecha no se han descrito variantes genéticas para este factor de transcripción en el contexto de estas enfermedades. Sin embargo, sí se han identificado polimorfismos en regiones no codificantes para otros factores de transcripción, como los SNPs de las regiones reguladoras de varias metaloproteasas (MMP). Las MMP, en el contexto de la EPOC, actúan degradando la matriz extracelular provocando la inflamación crónica de las vías aéreas. Por tanto, aquellas variantes genéticas que afectan directamente su expresión o función pueden impactar en la progresión de la enfermedad. Así, se han descrito SNPs en regiones

codificantes y no codificantes de *MMP2* (Price et al. 2001), en particular se ha identificado una transición de C/T en la posición -1306 del promotor que modifica la secuencia consenso del factor de transcripción Sp1 (CCACC). En este trabajo, se demuestra mediante ensayos reporteros y EMSA que la variación C(C/T)ACC provoca la disminución de la actividad promotora de *MMP2*, al destruir la secuencia consenso e impedir la unión del factor de transcripción Sp1. Por otro lado, se han analizado polimorfismos en *MMP12* identificando que el alelo menos frecuente G del SNP rs2276109 en su promotor se asocia un aumento de FEV1, el desarrollo de EPOC más tardío y con un menor riesgo de desarrollo de EPOC en fumadores (Hunninghake et al. 2009). Este alelo G ha sido descrito en otro trabajo como un polimorfismo que afecta la unión del factor de transcripción AP-1 disminuyendo la actividad promotora del gen de *MMP12* (Jormsjö et al. 2001), lo cual explicaría el efecto del SNP y su incidencia en un menor riesgo de padecer EPOC.

Por otro lado se ha identificado un SNP en las proximidades del gen que codifica para *hedgehog-interacting protein* (*HHIP*) asociados a EPOC, el alelo G del SNP rs13147758 se asocia a un aumento del cociente FEV1/FVC (OR: 0.397, pvalor 2.2E-06) pero no ha sido posible definir cuál es el gen o el efecto que explica dicha asociación (Wilk et al. 2009).

En relación a SAOS se ha identificado el alelo -460C localizado en una región no codificante del gen de *VEGF* asociado con un mayor riesgo de padecer SAOS (OR: 1,95; IC95%: 1,38– 2,76). Además se han identificado otros dos alelos que mapean en regiones codificantes de *VEGF* y se asocian a SAOS, pero no se describe cual es el efecto funcional que origina dichas asociaciones aunque se ha descrito que los niveles séricos de VEGF están elevados en pacientes con SAOS (Cao et al. 2014). Por otro lado, se ha encontrado una asociación significativa entre el alelo -308A en el promotor de *TNF- $\alpha$*  y SAOS (OR=1,8; IC95%: 1,18–2,75), lo cual podría contribuir a exacerbar la respuesta inflamatoria que presentan habitualmente estos pacientes (Riha et al. 2005).

Pese a que tomados en conjunto, todos estos trabajos identifican varias asociaciones entre variantes genéticas y diferentes aspectos de EPOC y SAOS, en muy pocos casos se puede demostrar causalidad entre la presencia del SNP y el fenotipo, poniendo de manifiesto el gran desafío en este tipo de estudios.

Recientemente, se realizó un estudio global denominado *Genotype-Tissue Expression* (GTEx), cuyo objetivo es analizar la asociación entre la presencia de variantes genéticas en el genoma y la expresión génica. En éste estudio multicéntrico se analiza la variación genética y el transcriptoma de diferentes tejidos humanos en miles de individuos, y sus resultados preliminares indican que los polimorfismos pueden tener un papel importante en la regulación de la expresión génica en función del tejido que se estudie y la presencia de patologías (The GTEx Consortium 2015). Estudios

anteriores, ya habían demostrado que entre un 5% y un 25% de los sitios de unión de factores de transcripción se ven alterados por la presencia de SNPs (Reddy et al. 2012; McVicker et al. 2013).

Todas estas evidencias, ponen de manifiesto el papel crucial de las variantes genéticas en regiones reguladoras no codificantes. Por lo tanto, la identificación de SNPs en sitios de unión de HIF en enfermedades que cursan con hipoxia, concretamente en EPOC y SAOS, es de gran interés ya que podrían explicar las diferencias fenotípicas observadas en los pacientes.

### **3 OBJETIVOS**





Las variantes genéticas en regiones no codificantes pueden alterar el patrón de expresión espacio-temporal y la magnitud de la regulación de la expresión génica. Previo a este trabajo de Tesis Doctoral, nuestro grupo demostró la existencia de SNPs que modifican la secuencia de los elementos de respuesta a hipoxia (HRE) alterando la regulación de la transcripción mediada por HIF, sugiriendo que la respuesta a la hipoxia en diferentes individuos es heterogénea y dependiente de las variantes genéticas interindividuales.

La enfermedad pulmonar obstructiva crónica (EPOC) y el síndrome de apnea obstructiva del sueño (SAOS) se caracterizan por el desarrollo de hipoxemia que da lugar a la activación del factor inducible por hipoxia (HIF), el cual parece jugar un papel central en la progresión de estas enfermedades y, en particular, en el desarrollo de comorbilidades. Por otro lado, la evolución de estos pacientes es heterogénea, pero el origen de esta variabilidad es desconocida.

En base a estos antecedentes, la hipótesis de partida de esta tesis es que la diferente progresión de los pacientes de EPOC y SAOS puede deberse, en parte, a la existencia de variantes genéticas interindividuales que localizan en sitios de unión de HIF alterando la respuesta génica a la hipoxia que se da en los tejidos de estos pacientes. Para poner a prueba esta hipótesis nos planteamos los siguientes objetivos:

1. Obtención de un listado fidedigno de regiones genómicas que unen HIF y contienen potenciales HREs funcionales.
2. Secuenciación de las regiones identificadas en el primer objetivo para identificar variantes en motivos RCGTG del genoma de pacientes con EPOC y SAOS.
3. Caracterización funcional de las variantes identificadas en el genoma de pacientes con EPOC y SAOS, respecto a su efecto en la transcripción génica mediada por HIF.
4. Análisis de correlación entre las variantes identificadas en motivos RCGTG y los fenotipos o comorbilidades de los pacientes con EPOC y SAOS.



## **4 MATERIALES Y MÉTODOS**





#### 4.1 Cultivos celulares y tratamiento de hipoxia

Las líneas celulares HeLa y HEK293T se crecieron en medio DMEM (Sigma). El medio de cultivo se suplementó con 50U/ml penicilina (Life Technologies), 50µg/ml estreptomicina (Life Technologies), 2,5 µg/ml fungizona (Sigma) y 10% (v/v) suero fetal bovino (Life Technologies). Las células se crecieron a 37°C en atmósfera húmeda al 5% de CO<sub>2</sub>. Las condiciones de hipoxia se establecieron utilizando una cámara Hypoxystation H35 (Don Whitley Scientific Limited) durante los tiempos indicados. La hipoxia se generó usando una mezcla de gas compuesta por 1% O<sub>2</sub>, 5% CO<sub>2</sub> y 94% N<sub>2</sub>, en atmósfera saturada de humedad, a una temperatura constante de 37°C.

#### 4.2 Generación de plásmidos

Las construcciones reporteras se generaron a partir de ADN genómico extraído de sangre periférica de pacientes con EPOC y SAOS utilizando Gentra Puregene Blood Kit (Qiagen, 158467). Se utilizaron individuos heterocigotos para las variantes de interés con el propósito de generar dos construcciones, una con el nucleótido de referencia (WT) y otra con la variante alternativa (SNP). Las regiones clonadas se indican en la [Tabla 1](#).

Plásmido receptor	Coordenadas genómica	Gen	Zona clonada	Plásmido resultante	Genotipo
pGL4.20	chrX:77359566:A	PGK1	chrX:77,359,301-77,359,819	PGK1 wt	CACGT
				PGK1 SNP HRE2	CACGA
Prol-	chr7:55254186	EGFR	chr7:55,252,975-55,255,081	EGFR wt	CACGT
				EGFR SNP	CACAT
	chr20:50366859	ATP9A	chr20:50,366,359-50,367,325	ATP9A wt	GCGTG
				ATP9A SNP	GCATG

**Tabla 1. Regiones reguladoras clonadas.** En la tabla se muestran las coordenadas de las regiones clonadas en el plásmido reportero pGL4.20 o Pro, el gen asociado a dicha región y el SNP identificado.

Los amplicones de cada región se clonaron, utilizando el kit In-fusion HD Cloning (Clontech), en el vector reportero *pGL4.20* (Promega). Las regiones promotoras (clasificadas según su ubicación respecto del cuerpo del gen) se clonaron entre los sitios XhoI y HindIII del vector pGL4.20. Por otro lado, para el estudio de regiones asociadas a actividad potenciadora de la transcripción (enhancers) se utilizó el vector *Prolactina* (Prol-) que contiene el gen de luciferasa bajo el control del promotor mínimo de prolactina de rata y por ello es de utilidad para el estudio de regiones con actividad de enhancer. Este vector reportero no es de origen comercial sino que ha sido generado y utilizado en el trabajo de Rincón et al. (Rincón & Flavell 1994) para estudiar la regulación transcripcional mediada por AP-1 en células T. En el contexto de la inducción de genes en hipoxia, el vector se utilizó por primera vez en un trabajo de Aragonés et al. (Aragones et al. 2001).

Las construcciones reporteras mutantes para cada uno de los HREs del gen de PGK1 se generaron usando el kit QuikChange® II XL Site-Directed Mutagenesis (Agilent, 200521-5), utilizando los cebadores indicados en la [Tabla 2](#).

Plásmido original	Cebador	Secuencia	Plásmido resultante
PGK1 wt	HRE1 Mut F	GGTTCGCGGCGTGCCGGGCTAGACAAACGGAAGCC	MUT HRE 1
	HRE1 Mut R	GGCTTCCGTTTGTCTAGCCCGGCACGCCGGAACC	
PGK1 wt	HRE2 Mut F	GCGAGGGTACTAGTGAGGCTAGCGGCTTCCGTTTGTCA	MUT HRE 2
	HRE2 Mut R	TGACAAACGGAAGCCGCTAGCCTCACTAGTACCCTCGC	
PGK1 wt	HRE3 Mut F	GCCTCCGGAGCGCTAGCCGGCAGTCGGCTCCC	MUT HRE 3
	HRE3 Mut R	GGGAGCCGACTGCCGGCTAGCGCTCCGGAGGC	

**Tabla 2. Primers utilizados para la mutagénesis de los HREs de PGK1.**

La identidad de todas las construcciones fue verificada por secuenciación de Sanger.

#### 4.3 Ensayos reporteros

Los ensayos reporteros fueron realizados en células HeLa que se sembraron 24 horas antes de la transfección en placas de seis pocillos (160.000 células / pocillo MW6). Se utilizó fosfato cálcico para transfectar una mezcla de 9 µg de ADN total por pocillo (compuesta por 3 µg del plásmido reportero indicado en cada caso, 0,6 µg de un plásmido que codifica para el enzima luciferasa-renilla bajo el control del promotor CMV (pRL-CMV Vector Promega) y 5,4 µg de pCDNA3). Pasadas ocho horas, se reemplazó el medio con la mezcla de transfección por medio fresco y 16 horas después se sembraron las células transfectadas en placas de 24 pocillos, realizando para cada condición experimental triplicados. Una vez adheridas, las células fueron expuestas a condiciones de normoxia (21%O<sub>2</sub>) o hipoxia (1%O<sub>2</sub>). Finalizado el tratamiento, se lisaron las células para analizar las actividades luciferasa y renilla usando el *Dual Luciferase Reporter Assay System* (Promega), según las indicaciones del fabricante, en un luminómetro *Lumat LB9507* (Berthold Technologies). La actividad luciferasa se normalizó por la actividad renilla, para corregir la eficiencia de transfección entre muestras y se calculó las medias y desviaciones estándar. Se realizaron tres réplicas para cada condición experimental y al menos cuatro experimentos independientes para cada una de las regiones analizadas.

#### 4.4 Selección de pacientes

El reclutamiento de los pacientes se realizó en las consultas de los Servicios de Neumología del Hospital Universitario La Paz y Hospital Universitario La Princesa, ambos en Madrid.

Todos los pacientes que forman parte de este estudio tienen una edad comprendida entre 35 y 80 años y han dado su consentimiento informado y el estudio ha sido aprobado por los comités de ética de la investigación clínica de los dos centros participantes. Las muestras han sido anonimizadas y se analizaron datos agregados sin que se pueda rescatar la identidad de los pacientes.

En total, se reclutaron 98 pacientes con EPOC y 144 pacientes con SAOS.

#### Sujetos de estudio con EPOC:

Se realizó un reclutamiento consecutivo de pacientes diagnosticados de EPOC moderada-muy grave ( $FEV_1/FVC$  post-broncodilatador  $< 0,70$  y  $FEV_1$  post-broncodilatador  $< 80\%$  pred.) (Global Strategy for the Diagnosis, Management and Prevention of COPD), clínicamente estables durante al menos ocho semanas, con tratamiento óptimo según la normativa de la *American Thoracic Society/European Respiratory Society* (ATS/ERS) (B.R. Celli et al. 2004) no modificado en las últimas ocho semanas y con historia de tabaquismo ( $> 10$  paquetes x año). Se consideraron criterios de exclusión: diagnóstico de asma bronquial, enfermedad pulmonar intersticial difusa, enfermedad de la caja torácica o neuromuscular, neoplasia pulmonar o cualquier situación con esperanza de vida menor de 1 año.

Valoración clínica-funcional: en todos los casos se realizó una evaluación antropométrica y clínico-funcional, que incluía anamnesis y exploración clínica; puntuación de la disnea basal en la escala modificada del *Medical Research Council* (mMRC) (Mahler 1987); calidad de vida relacionada con la salud mediante el cuestionario respiratorio St George (SGRQ) (Ferrer et al. 1996); espirometría, pletismografía, capacidad de difusión de CO y prueba de la caminata de seis minutos, según las normativas de la ERS/ATS (Miller et al. 2005; Wanger et al. 2005; MacIntyre et al. 2005; Quanjer et al. 1993; Crapo et al. 2002); hemograma, perfil bioquímico y proteína C reactiva; electrocardiograma y ecocardiografía. Según estos datos, los pacientes se clasificaron en los siguientes grupos de comorbilidad:

- Hipertensión pulmonar, cuando la presión sistólica en la arteria pulmonar fue superior a 35 mmHg (Proceedings of the 4th World Symposium on Pulmonary Hypertension, February 2008, Dana Point, California 2009).
- Hipertensión arterial, definida por una presión arterial por triple toma superior a 140/90 mmHg, por monitorización ambulatoria de presión arterial de 24 horas mayor de 135/85 mmHg o por la existencia de un diagnóstico previo de hipertensión arterial asociado a la toma de medicación antihipertensiva (Mancia et al. 2007).
- Diagnóstico clínico de cualquier tipo de neoplasia, con confirmación histológica.

- Diagnóstico de cardiopatía isquémica, según los criterios de la *European Society of Cardiology/American College of Cardiology* (The Joint European Society of Cardiology American College of Cardiology Committee 2000).
- Riesgo cardiovascular, definido por la presencia de hipertensión arterial, diabetes, dislipidemia o síndrome metabólico. El diagnóstico de diabetes mellitus se estableció cuando un paciente estaba recibiendo tratamiento con insulina y/o antidiabéticos orales o presentaba una glucemia basal superior a 125 mg/dl (American Diabetes Association. 2013). Se consideró que tenía dislipidemia cuando refería dicho diagnóstico y el consumo de fármacos hipolipemiantes o presentaba un perfil lipídico alterado (colesterol total  $\geq$  200 mg/dL, triglicéridos  $\geq$  180 mg/dL, HDL-colesterol  $\leq$  40 mg/dL, o LDL-colesterol  $\geq$  150 mg/dL) (Bartels et al. 2007). A su vez, el síndrome metabólico fue definido según los criterios establecidos por el *National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults* (Grundy et al. 2004).

Además, la comorbilidad fue valorada globalmente mediante la escala específica para la EPOC COTE (Divo et al. 2012).

Por último, la gravedad de la EPOC se clasificó en función del fenotipo clínico según la guía GesEPOC (diferenciando pacientes agudizadores de no agudizadores, por la presencia de más de dos exacerbaciones moderadas-grave al año) (Miravittles et al. 2014), grado de limitación al flujo aéreo, grupos de riesgo GOLD (Global Strategy for the Diagnosis, Management and Prevention of COPD) y mediante la escala multidimensional BODE (Bartolomé R Celli et al. 2004).

Se obtuvieron muestras de sangre periférica de todos los sujetos destinada a la obtención de ADN genómico para la determinación de polimorfismos.

#### Sujetos de estudio con SAOS:

Según los criterios del *Spanish Sleep Network*, se consideró que un individuo tiene SAOS ante la presencia de síntomas sugestivos de la enfermedad (sobre todo, somnolencia con una puntuación en la escala de Epworth mayor de 10), un índice de apneas-hipopneas (IAH)  $> 5 \text{ h}^{-1}$  (detectado por poligrafía respiratoria o videopolisomnografía vigilada) y al menos un 80% de sus apneas-hipopneas eran de tipo obstructivo (Lloberes et al. 2011).

Se consideraron como criterios de exclusión: hipertensión arterial grave ( $>180/120 \text{ mmHg}$ ), hipertensión arterial secundaria conocida, somnolencia diurna excesiva muy grave, ocupación de alto riesgo (conductor profesional, manipulador maquinaria pesada, etc.), diagnóstico de asma bronquial, bronquiectasias, cáncer de pulmón, enfermedad pulmonar restrictiva, trastornos de la caja torácica o cirugía torácica previa, enfermedad neuromuscular o anomalía de la función tiroidea, obesidad



mórbida ( $BMI > 40 \text{ Kg/m}^2$ ), infección respiratoria en los últimos dos meses, o entorno sociofamiliar inadecuado que, a juicio de los investigadores, pudiera limitar el cumplimiento terapéutico y participación en los procedimientos del estudio.

Todos los pacientes se sometieron a una evaluación similar a la descrita en los pacientes con EPOC, identificándose los siguientes subgrupos, según los criterios ya referidos:

- Hipertensión pulmonar
- Síndrome metabólico
- Hipertensión arterial
- Cáncer
- Dislipidemia
- Diabetes mellitus
- Cardiopatía isquémica.

Además, en los pacientes con SAHS, se establecieron estos subgrupos adicionales:

- Troboembolismo pulmonar, diagnosticado mediante angioTC, según las guías de la European Society of Cardiology (Torbicki et al. 2008).
- Insuficiencia cardíaca diagnosticada por la presencia de clínica compatible y la evidencia ecocardiográfica de disfunción sistólica o diastólica del ventrículo izquierdo (Redfield et al. 2003).
- Depresión/ansiedad. Evidenciada por un diagnóstico previo y la utilización de medicación específica (antidepresivos o ansiolíticos) o por una puntuación alterada en los cuestionarios *Beck depression inventory* (Beck et al. 1961) o *State-Trait Anxiety Inventory* (Spielberger et al. 1970).

Se obtuvieron muestras de sangre periférica de todos los sujetos destinada a la obtención de ADN genómico para la determinación de polimorfismos.

#### 4.5 Extracción de ARN y qPCR-RT

Los niveles de ARNm se determinaron mediante ensayos de PCR cuantitativa a tiempo real (qPCR-RT) mediante el uso de cebadores específicos o de sondas *TaqMan*. El ARN total se extrajo con el kit RNeasy Mini (Qiagen, 74106) y se trató con RNase-Free DNase Set (Qiagen, 79254). Para la síntesis de ADNc se retrotranscribió  $1\mu\text{g}$  del ARN total de cada muestra (Transcriptor First Strand cDNA Synthesis kit, Roche). Para las reacciones de amplificación con cebadores específicos se utilizaron  $5 \mu\text{l}$  de una dilución 1:20 del ADNc y Power SYBR green PCR Master Mix (Applied Biosystems, 4367659).

Para la detección del ARNm de PGK1 y RUVBL2 se utilizaron sondas *TaqMan* junto con Taqman Universal Master Mix II, no UNG (Applied Biosystems, 4440040), según las instrucciones el fabricante. En los ensayos con sondas *TaqMan*, la síntesis de ADNc se realizó como se explicó anteriormente. A

continuación, se diluyó el ADNc hasta una concentración final de 2,5 ng/μl y se llevó a cabo la reacción con 2 μl de producto de ADNc, 0,5 μl sonda *TaqMan* 20x, 5 μl *Master mix* 2x (Applied Biosystems) y 2,5 μl de agua destilada. Las sondas *TaqMan* utilizadas fueron: *PGK1* (Hs00943178\_g1) y *RUVBL2* (Hs00272632\_m1).

Se utilizó, independientemente de cómo se obtuvo el ADNc, un termociclador StepOne Realtime PCR System (Applied Biosystems), y los resultados se analizaron con el programa Step One. Los niveles de expresión se calcularon utilizando el método  $\Delta\Delta C_t$  utilizando  $\beta$ -actina como referencia. Los datos fueron analizados por el método de cuantificación relativa por compensación de  $C_t$  (threshold cycle, umbral de detección establecido). En este método se calcula la media entre los triplicados, y los niveles de expresión del gen de interés se normalizan por la media de los triplicados del gen constitutivo utilizado ( $\beta$ -actina). A continuación, la cantidad de ARNm del gen de interés de una muestra determinada (m) se expresa en relación a los niveles de dicho ARNm en una muestra calibradora (c), de tal modo que la cantidad relativa de un gen de interés en una muestra m es igual a  $2^{-\Delta\Delta C_t}$ , donde  $\Delta\Delta C_t = (\Delta C_t)_m - (\Delta C_t)_c$ , siendo  $\Delta C_t = C_t$  gen de interés –  $C_t$  gen constitutivo. Todos los resultados fueron confirmados en al menos tres experimentos independientes. En la [Tabla 3](#) se indican los cebadores utilizados para estos experimentos.

Gen	Cebador	Secuencia
<i>EGFR</i>	EGFR D	TAGACAAACGGAAGGGTTCGCGGCGTGCCGGGCC
	EGFR R	GGCTTCGTTTGTCTAGCCCGGCACGCCGGAACC
<i>EGLN3</i>	EGLN3 D	GGTTCGCGGCGTGCCGGGCTAGACAAACGGAAGCC
	EGLN3 R	TGACAAACGGAAGCCGCTAGCCTCACTAGTACCCTCGC
<i><math>\beta</math>actina</i>	ACT D	CGGCGTGCCGGGCTAGGTTCCGACAAACGGAAGCC
	ACT R	GGGAGCCGACTGCCGGCTAGCGCTCCGGAGGC

**Tabla 3. Primers utilizados para la detección del ARNm en qPCR-RT.** Se indica el locus en estudio y la secuencias del cebador Directo (D) y cebador Reverso (R).

#### 4.6 Perfil de expresión génica de alta resolución

Para determinar la regulación transcripcional de EGFR en hipoxia se realizó un ensayo de perfil de expresión génica de alta resolución como se describe en el trabajo de Dolken *et. al.* (Dölken et al. 2008). Este ensayo se basa en el marcaje metabólico de ARN de nueva síntesis utilizando 4-tiouridina (marcaje 4sU), un nucleótido de uridina modificado con un grupo tiol, que permite detectar transcritos recientemente sintetizados sin interferir en el crecimiento celular y la expresión génica. Para este experimento se expusieron células HeLa a 21% o 1% de oxígeno durante 8 horas y en las últimas 2 horas de tratamiento se realizó un pulso de 400mM de 4sU (Sigma, T4509) para permitir la incorporación de 4sU- trifosfato en el ARN de nueva síntesis. Después de aislar el ARN total, la

fracción de ARN marcada con 4sU fue específicamente biotinilada mediante la generación de un enlace disulfuro entre la biotina y el grupo tiol del 4sU, de manera que el ARN total puede ser separado con una alta pureza en ARN marcado (ARN de nueva síntesis) y en ARN no marcado (ARN pre-existente) utilizando bolas magnéticas recubiertas con estreptavidina ( $\mu$ Mac<sup>s</sup> Streptavidin Kit; Miltenyi, 130-074-101). Finalmente, el ARN marcado se recupera de estas bolas magnéticas mediante la adición de un agente reductor, DL-Ditiotreitol 100 mM (Sigma, 43815), que rompe el enlace disulfuro y libera el ARN de nueva síntesis. Finalmente el ARN de nueva síntesis se purificó utilizando Ultrapure™ Phenol: Chloroform: Isoamylalcohol (Invitrogen, 15593-031), y el ARN pre-existente se purificó usando RNeasy Mini Kit (QIAGEN, 74106). Las muestras de cada condición fueron utilizadas para qPCR-RT.

#### **4.7 Análisis de proteínas**

##### **i) Extracción de proteínas celulares**

Los extractos de proteínas celulares se obtuvieron homogeneizando las células, tras dos lavados con PBS, con tampón RIPA (50 mM Tris-HCl pH 8, 150 mM NaCl, 0,02% azida sódica, 0,1% SDS, 1% NP-40, 1% deoxicolato sódico) conteniendo inhibidores de proteasas (Complete ULTRA tablet, Roche 06538304001). La lisis se realizó en hielo durante 15 min, tras lo cual se usaron agujas de 23 G para fragmentar el ADN y se centrifugaron a 12000 rpm a 4º C durante 15 min para descartar los restos celulares. La cuantificación de proteínas se llevó a cabo mediante el kit BioRad DC protein assay (BioRad, 5000112) compatible con detergentes y basado en el método Lowry. Finalmente, se añadió tampón Laemli a los extractos, se hirvieron 5 min a 100º C y se almacenaron a - 80º C.

##### **ii) Inmunodetección de proteínas mediante Western blot**

Las muestras de proteínas se separaron en geles de 10% poliacrilamida-SDS (PAGE-SDS), en los que se cargaron 20  $\mu$ g de proteína por cada carril del gel. A continuación, los geles se transfirieron a membranas de PVDF (Immobilon®-P, Millipore, IPVH100010) mediante transferencia líquida (20% metanol, 39 mM glicina y 48 mM Tris) a 100 V a 4º C durante 2 h aproximadamente. Las membranas se incubaron en tampón de bloqueo (0,1% Tween-20, 5% leche en polvo desnatada en TBS) a temperatura ambiente durante 1 h, y posteriormente se incubaron con los anticuerpos primarios correspondientes ([Tabla 4](#)) diluidos en tampón de bloqueo a 4º C toda la noche. Al día siguiente, tras lavar tres veces con 0,1% Tween-20 en TBS, se incubaron con el correspondiente anticuerpo secundario conjugado con peroxidasa ([Tabla 4](#)) diluido en tampón de bloqueo durante 1 h a temperatura ambiente. Por último, se lavaron tres veces las membranas, y la detección de las bandas se realizó mediante reacción de quimioluminiscencia con el reactivo ECL (Amersham ECL Western Blotting Detection Kits, GE Healthcare, RPN2108) utilizando películas autorradiográficas (Agfa, ENKMOV).

Anticuerpos primarios	Especie	Procedencia	Referencia	Dilución
HIF1 $\alpha$ humano	Ratón	Abcam	Ab 2185	1:1000
EGFR	Ratón	BD Transduction Laboratories	610017	1:1000
$\beta$ Lamina	Cabra	Santa Cruz	sc-6216	1:1000
Anticuerpos secundarios	Especie	Procedencia	Referencia	Dilución
Anti- Ratón	Cabra	Promega	W402B	1:5000
Anti-Cabra	Burro	Santa Cruz	sc-2020	1:5000

**Tabla 4. Anticuerpos empleados para Western blot (WB).** Se indican los anticuerpos primarios y secundarios utilizados para WB.

#### 4.8 Determinación de regiones reguladoras que unen HIF en el genoma humano

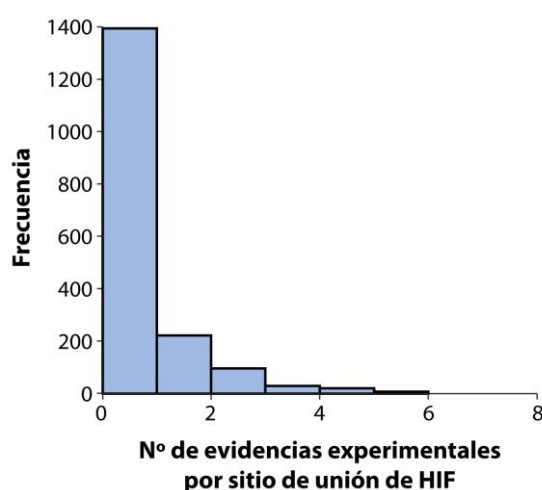
Los sitios de unión de HIF a nivel genómico se obtuvieron de la información publicada en 6 estudios independientes en los que la unión de HIF se determinó mediante inmunoprecipitación de cromatina seguida de hibridación a microarrays genómicos (ChIP-chip) o secuenciación masiva (ChIP-seq), las características de cada uno de estos estudios se indican en la [tabla 5](#). En uno de los estudios se determinó tanto el patrón de unión de HIF1- $\alpha$  como el de EPAS1, por lo que, finalmente, obtuvimos 7 listados de regiones de unión de HIF. En los casos en los que las coordenadas de unión de HIF utilizaban como referencia el ensamblado hg18, fueron convertidas a GRCh37/hg19 utilizando el programa LiftOver del UCSC.

Publicación	Técnica	Anticuerpo	Tipo celular	Tratamiento
Villar, 2012 Plos One	ChIP-ChIP	HIF-1 $\alpha$	HeLa	6 horas, 1% O <sub>2</sub>
Xia, 2009 Genome Biology	ChIP-ChIP	HIF-1 $\alpha$	U87	4 horas ,0,5% O <sub>2</sub>
Mole, 2009 JBC	ChIP-ChIP	HIF-1 $\alpha$	MCF7	16 horas, 2 mM DMOG
Xia, 2009 PNAS	ChIP-ChIP	HIF-1 $\alpha$	HepG2	4 horas, 0,5 %O <sub>2</sub> .
Schödel, 2011 BLOOD	ChIP-Seq	HIF-1 $\alpha$ EPAS1	MCF7	16 horas, 0.5% O <sub>2</sub> ó 2mM DMOG
Schödel 2012 Nature Genetics	ChIP-Seq	EPAS1	786-O	1% O <sub>2</sub> for 16 h ó 1 mM DMOG

**Tabla 5. Experimentos de ChIP-chip y ChIP-Seq seleccionados a partir de 6 publicaciones independientes.** Se muestran los siete experimentos de ChIP-chip y ChIP-Seq seleccionados a partir de experimentos independientes para la determinación de las regiones de unión de HIF en el genoma. En la tabla se indica la información relativa a la publicación, la técnica realizada, la isoforma de HIF para la que se realizó el ChIP, la línea celular, y la duración y tipo de tratamiento.

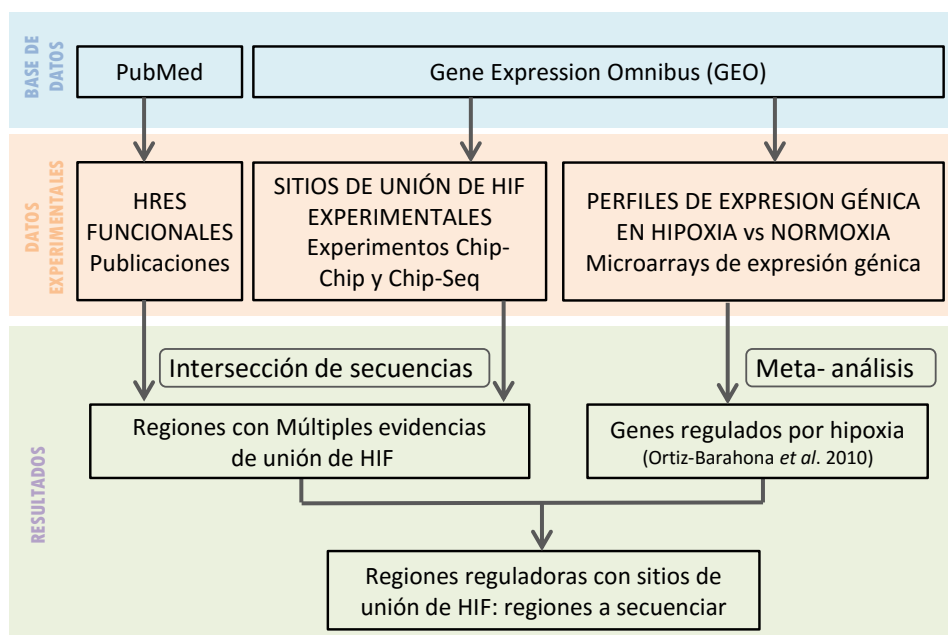
Además, recopilamos un listado de elementos de respuesta a hipoxia (“Hypoxia Response Elements”, HREs) descritos en la bibliografía como resultado del estudio de la inducción de genes

específicos en respuesta a hipoxia. En este análisis solo se incluyeron los HREs que se habían descrito de forma rigurosa, resultando en un total de 57. Las regiones de unión de HIF que representan este listado de HREs, se incorporó a las regiones identificadas en análisis globales (ANEXO I) como una evidencia experimental más. Los sitios de unión identificados en todos estos estudios se combinaron en una única tabla que recogía, para cada región, el número de evidencias experimentales que coincidían en esta región. En la mayoría de los casos la región se encontraba unida por HIF en un único estudio (Figura 6). Para la generación de estas tablas escribimos un programa (“script”) específico en el lenguaje de programación Python.



**Figura 6. Evidencias experimentales por sitios de unión de HIF.** Se indica el número de evidencias experimentales para una misma región de unión de HIF.

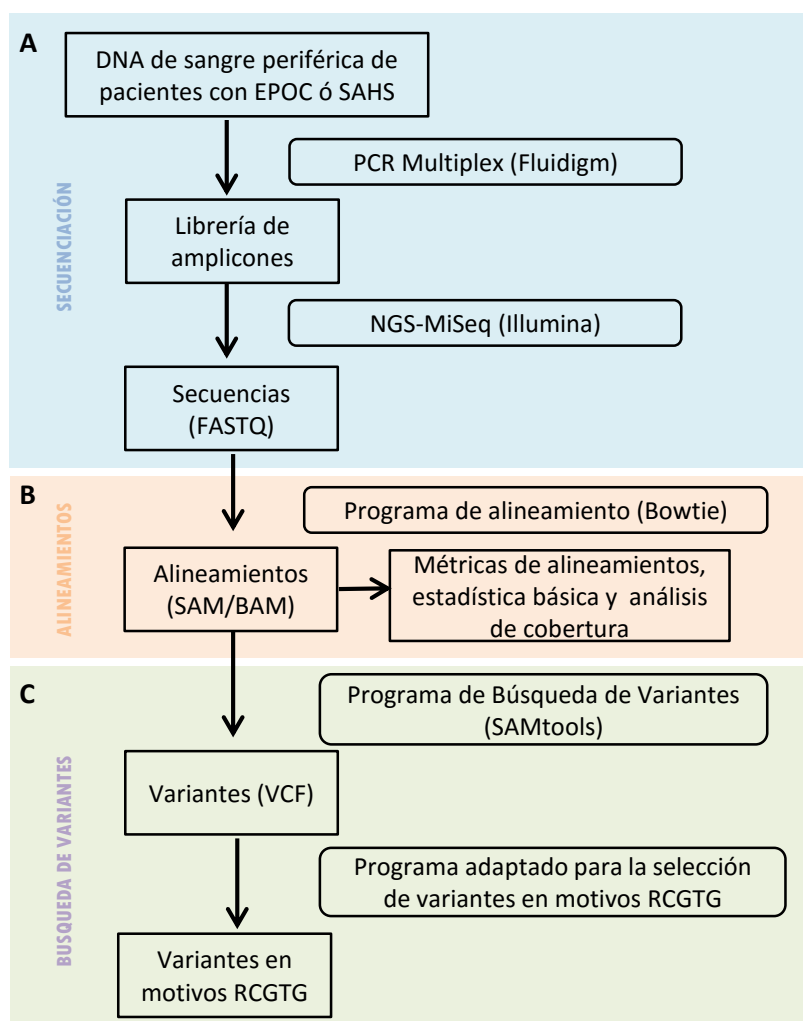
En la Figura 7 se muestra la estrategia utilizada para definir las regiones a secuenciar que poseen sitios de unión de HIF. En primer lugar, una vez recopilados y analizados los datos experimentales se integraron los HREs funcionales con las regiones que contienen sitios de unión de HIF, determinando regiones con múltiples evidencias de unión de HIF. A continuación, estas regiones se asociaron a genes regulados por hipoxia y se definió el listado de regiones en zonas reguladoras con alta probabilidad de contener HREs funcionales.



**Figura 7. Esquema de la estrategia utilizada para definir las regiones a secuenciar que poseen sitios de unión de HIF.** La estrategia utilizada se basa en la integración de tres tipos de datos experimentales: i) HRES funcionales obtenidos de artículos (PubMed), ii) sitios de unión de HIF obtenidos de experimentos de ChiP-chip y ChiP-seq y iii) perfiles de expresión génica en hipoxia versus normoxia obtenidos de la base de datos Gene Expression Omnibus (GEO), a partir de los que se determinaron los sitios de unión de HIF experimentales y perfiles de expresión de génica en hipoxia. Al integrar los HRES funcionales con las regiones que contienen sitios de unión de HIF se determinaron regiones con múltiples evidencias de unión de HIF. Por último, estas regiones se asociaron a genes regulados por hipoxia y se definió el listado de regiones en zonas reguladoras con alta probabilidad de contener HRES funcionales, que fueron seleccionadas para secuenciar en los pacientes de EPOC y SAOS.

#### 4.9 Secuenciación masiva

Para la preparación de la librería de amplicones, se extrajo ADN genómico con Gentra Puregene Blood Kit (Qiagen, 158467) a partir de sangre periférica de los pacientes incluidos en el estudio y las regiones de interés se amplificaron mediante PCR *Multiplex* utilizando cebadores diseñados por Fluidigm. Finalmente, las librerías generadas se secuenciaron en la plataforma MiSeq de Illumina (Figura 8A).



**Figura 8. Esquema de secuenciación y búsqueda de variantes genéticas en motivos RCGTG.** Este proceso se divide en tres etapas; **A)** Secuenciación; **B)** Alineamientos; **C)** Búsqueda de variantes. La secuenciación (**A**) consta en primer lugar de la preparación de la muestra, a partir de la extracción de ADN de sangre periférica de pacientes con EPOC o SAHS, seguida de la preparación de la librería de las regiones mediante PCR multiplex que finalmente serán secuenciadas por técnicas de secuenciación masiva. Una vez obtenidas las lecturas se realizan el alineamiento (**B**) de las secuencias al genoma de referencia, a la vez que se realizan estadísticas y análisis de la secuenciación y del alineamiento. Finalmente se realiza la búsqueda de variantes (**C**) en primera instancia se obtienen todas las variantes presentes en las regiones secuenciadas y por último se identifican las variantes en motivos RCGTG.

Las lecturas resultantes de la secuenciación de cada muestra (en formato FASTQ) se alinearon al genoma de referencia (versión GRCh37/hg19), utilizando el programa Bowtie (versión 2.1.0) (Langmead et al. 2009) con los valores por defecto de los argumentos (Figura 8B). Los alineamientos resultantes (formato SAM/BAM) se utilizaron para identificar variantes con respecto al genoma humano de referencia. Para ello utilizamos SAMtools (versión 0.1.19) y bcftools (versión 0.1.19) (Li et al. 2009; Li 2011)

Por último, las variantes genéticas presentes en las regiones secuenciadas se analizaron mediante un programa desarrollado en el laboratorio escrito en Python, que determina si las variantes se encuentran en un motivo RCGTG del genoma o crean nuevos motivos RCGTG (Figura 8C).

#### 4.10 Edición del genoma mediante el sistema CRISPR/Cas9

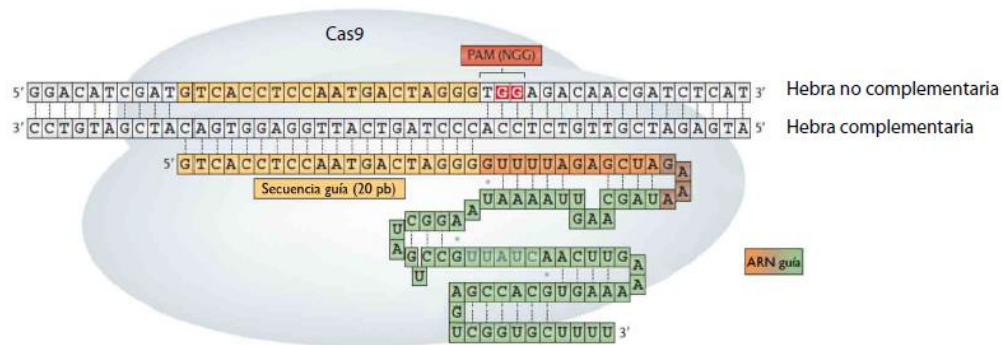
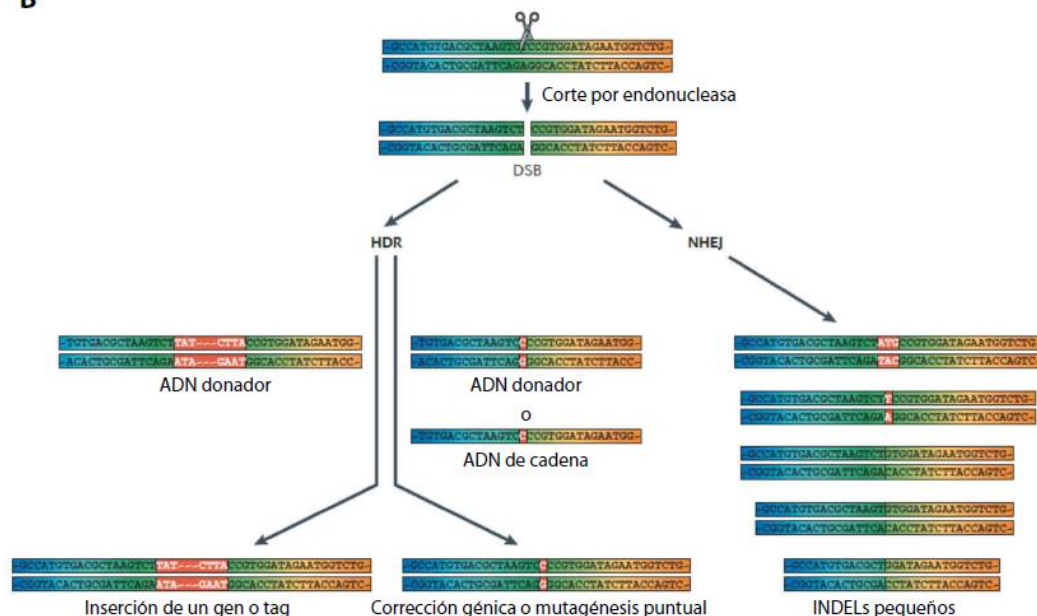
El sistema CRISPR/Cas9 más utilizado en la actualidad está basado en el sistema inmune adaptativo de *S.pyogenes*. Se compone de tres elementos principales: la endonucleasa Cas9, una secuencia de ADN diana que contiene el motivo adyacente protoespaciador (PAM, de sus siglas en inglés Protospacer Adjacent Motif) (NGG) y un ARN guía. Este último consta de dos partes, en el extremo 5' se encuentran 20 nucleótidos complementarios a la secuencia diana que se desea editar, y en el extremo 3' se encuentra la secuencia de ARN implicada en el mantenimiento de la estructura (ARN andamio) (Figura 9A).

En primer lugar el ARN guía y la endonucleasa Cas9 forman un complejo, que a través de Cas9 es capaz de reconocer la PAM presente en el ADN en las proximidades del sitio a editar (Carroll 2014) y a continuación el ARN guía se une por complementariedad de bases al locus de ADN a editar, activando la endonucleasa Cas9 que corta ambas hebras del ADN 3 bases en dirección 5' de la secuencia PAM (Jiang et al. 2015). Estos cortes de doble cadena pueden resolverse por recombinación homóloga (HDR) en presencia de un ADN donador, que contenga en su secuencia el cambio que se desea introducir, dando lugar a la edición precisa del genoma. Para ello puede utilizarse ADN donador de doble cadena o ADN de simple cadena, dando lugar a la inserción o sustitución de secuencias de un único nucleótido hasta inserciones de un gen completo (Knock in). Por otro lado estos cortes de doble cadena pueden repararse por la unión de extremos no homólogos (NHEJ), la cual no requiere de ADN donador y por tanto genera pequeñas inserciones y deleciones (INDELs), de +3 hasta -10 bases nucleotídicas (Figura 9B).

En este trabajo se utilizó el sistema CRISPR/Cas9 para editar el genoma de la línea celular HEK293T en tres locus: a) *EGFR*: introduciendo el SNP Chr7:55.254.186:A en un intrón de dicho gen; b) *EGLN3*: mutando el HRE descrito de *EGLN3* chr14:34.407.696-34.407.700 modificando ACGTG a **GCTAG**; c) *PGK1*: mutando 2 HREs funcionales del promotor de *PGK1*, HRE1 chrX:77.359.543-77.359.547 (ACGTG/**GGATC**) y HRE2 chrX:77.359.562-77.359.566 (CACGT/**CATGG**).

*EGLN3* se uso como un control de la edición, mutando un HRE que se ha caracterizado previamente en el laboratorio (Pescador et al. 2005).



**A****B**

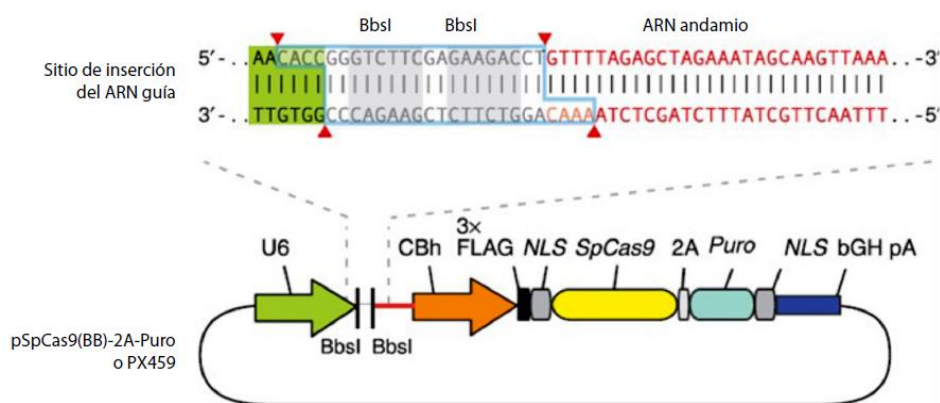
**Figura 9. Esquema del sistema de edición genómica CRISPR/Cas9.** A) Representación esquemática del complejo, la nucleasa Cas9-ADN diana-ARN guía. En amarillo se indica la secuencia diana que coincide con la secuencia guía, y en verde el ARN andamio. En rojo se indica la secuencia PAM (NGG) en el genoma a editar. B) Edición mediante endonucleasa Cas9. La endonucleasa Cas9 realiza un corte de doble cadena en el ADN diana 3 bases en dirección 5' de la secuencia NGG. El corte de doble cadena puede ser reparado mediante 2 vías: i. Recombinación homóloga (HDR) en presencia de ADN donador dando lugar a la edición precisa del genoma, insertando un gen completo o realizando un cambio en una única base; ii) Reparación por la unión de extremos no homólogos (NHEJ), generando pequeñas inserciones y deleciones (INDELs) (Figura adaptada de (Kim & Kim 2014) Hyongbum Kim et al, 2014 poner la CITA)

### i) CRISPR/Cas9: Generación de plásmidos.

Una vez determinados los locus a editar se diseñaron 4 ARN guías para cada uno (Tabla 6). El diseño de estas secuencias se realizó con la herramienta informática desarrollada por el grupo de Zhang *CRISPR Design Tool* (<http://tools.genome-engineering.org>) y se seleccionaron aquellas secuencias que fuesen más próximas al sitio a editar. Estas secuencias se clonaron en el plásmido PX459 (#48139 Addgene) entre los sitios de corte para BbsI, de manera que el ARN guía se insertó en el extremo 5' del ARN de andamiaje (Figura 10). Este plásmido contiene por un lado el ARN guía bajo el promotor U6 y por otro la endonucleasa Cas9 de *S. pyogenes* y el gen de resistencia a Puromicina bajo el promotor CBh.

Locus	ARN Guía	Secuencia ARN Guía
PGK1	SgARN 1	CCGGCACGCCGCGAACCGCAAGG
	SgARN 2	GGCGTGCCGGACGTGACAAACGG
	SgARN 3	CGGCTTCCGTTTGTACGTCCGG
	SgARN 4	GAGGGTACTAGTGAGACGTGCGG
EGFR	SgARN 1	GCTGTGTGATGTGAACGTGCTGG
	SgARN 2	GCTGGGTAAGTCATCCTCTCTGG
	SgARN 3	ACACAGCTTTGTGGATTCTTAGG
	SgARN 4	GTTACATCACACAGCTTTGTGG
EGLN3	SgARN 1	GGGCTACGTGCGCTGCGTGAGGG
	SgARN 2	CTACGTGCGCTGCGTAGGGTGG
	SgARN 3	CTGGCGGTGTGCTCGGGTGTGGG
	SgARN 4	AGGGTGATGGCTGAGTGACAGGG

**Tabla 6. ARN guías utilizados para edición del genoma mediante CRISPR/Cas9.** Se indica la secuencia de los ARN guías seleccionados para editar el genoma mediante el sistema CRISPR/Cas9 en regiones no codificantes asociadas a *PGK1*, *EGFR* y *EGLN3*.



**Figura 10. Esquema del plásmido PX459, utilizado para la edición del genoma mediante el sistema CRISPR/Cas9.** Se indican los sitios de corte para la enzima de restricción BbsI donde se han clonado los ARNs guías en el plásmido PX459 bajo el control del promotor U6. Además se indica el promotor CBh que regula la expresión de la enzima Cas9 de *S. pyogenes*, así como el gen de resistencia a Puromicina. (Figura adaptada de (Ran et al. 2013)).

## ii) CRISPR/Cas9: Condiciones de transfección.

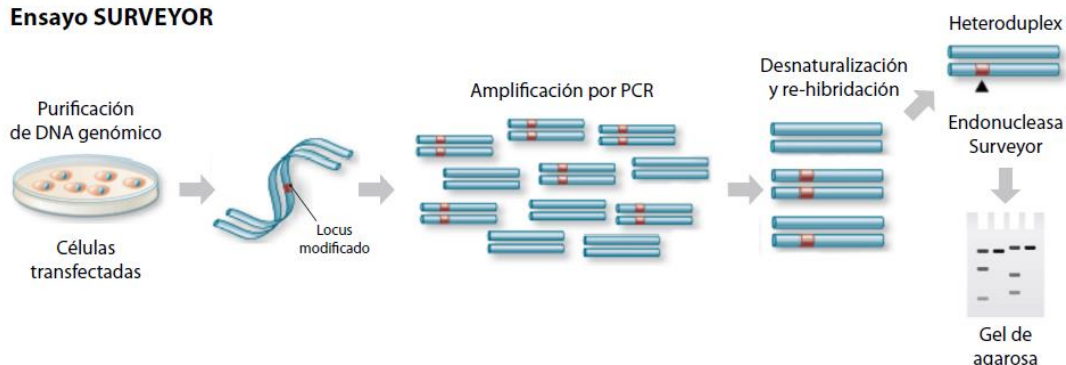
Para realizar el ensayo Surveyor se transfectaron 230.000 células HEK293T en placas de 24 pocillos utilizando 0,5µg del plásmido PX459 conteniendo uno de los 4 ARN guías para cada locus, en presencia de 1,25ul de Lipofectamine® 2000 Transfection Reagent (Thermo Fisher Scientific, cat. No. 11668019) en 100ul de OptiMEM (Gibco) durante 16 horas y tras 24 horas de transcurrida la transfección se extrajo el ADN.

Para la edición génica de cada uno de los locus se transfectaron 1,2 millones de células HEK293T en placas de 6 pocillos con una mezcla de ADN que contiene: 2,5 µg de plásmido que tiene la secuencia de ARN guía más eficiente para cada locus y 0,6 µg de ADN donador de cadena simple, indicados en la [tabla 6](#), en presencia de 6,25ul de Lipofectamine® 2000 Transfection Reagent (Thermo Fisher Scientific, cat. No. 11668019) en 500ul de OptiMEM (Gibco) durante 16 horas. Tras 48 horas de transcurrida la transfección se aislaron clones. El aislamiento de clones se realizó por dilución límite, plaqueando 0,5 células en placas de 96 pocillos. Se identificó la presencia de clones en el microscopio y se crecieron durante 15 días, y se genotiparon por secuenciación de Sanger para determinar la edición del genoma ocurrida en cada caso.

Aunque el plásmido PX459 contiene selección de puomicina, en ninguna de las transfecciones se realizó la selección con este antibiótico ya que se comprobó transfectando un plásmido de expresión con GFP que la eficiencia de transfección de las HEK293T es muy elevada (superior al 90%). Además, posteriormente se publicó que este plásmido PX459 contiene una mutación puntual en el gen de Puomicina que hace que dicha resistencia sea menos efectiva en algunas líneas celulares. Para todas las transfecciones se utilizó Lipofectamine® 2000 Transfection Reagent (Thermo Fisher Scientific, cat. No. 11668019)

## iii) CRISPR/Cas9: Ensayo Surveyor

Una vez clonados los ARNs guías para cada locus se realizó el ensayo Surveyor (Qiu et al. 2004). Este ensayo se basa en la actividad catalítica de la endonucleasa Surveyor que reconoce y corta en bases desapareadas, ya sea por la presencia de polimorfismos de nucleótido único, inserciones pequeñas o deleciones de bases. Este ensayo involucra cuatro pasos principales: i. Amplificación del ADN por PCR; ii. Desnaturalización y posterior hibridación del producto de PCR para la formación de heteroduplex; iii. Tratamiento del ADN re-hibridado con la endonucleasa Surveyor (Transgenomic, cat. no. 706025) para digerir los heteroduplex; y iv. Separación y detección de los productos de ADN digeridos en un gel de agarosa ([Figura 11](#)).

**Ensayo SURVEYOR**

**Figura 11. Representación esquemática del ensayo Surveyor.** Este ensayo es una herramienta para detectar mutaciones y polimorfismos en el ADN. Se basa en amplificar por PCR la región en estudio, seguida de un ciclo de desnaturalización y re-hibridación de los productos de PCR amplificados. Finalmente los productos re-hibridados son digeridos con la endonucleasa Surveyor que reconoce y corta en hebras desapareadas y el resultado se observa en un gel de agarosa.

Para realizar este ensayo y así poder determinar el ARN guía con el cual se logra la mayor porcentaje de células editadas, se transfectaron HEK293T y tras 24 horas se extrae el ADN. Una vez extraído el ADN se utilizan los cebadores indicados en la [Tabla 7](#) se realizó una PCR de las regiones de interés, y posteriormente se realiza el ensayo Surveyor siguiendo las instrucciones de la publicación de Ran, F Ann et al (Ran et al. 2013). Por último, se calculó el porcentaje de INDELS cuantificando la intensidad de las bandas digeridas y sin digerir según la formula indicada en trabajo de Pyzocha et al (Pyzocha Neena et al. 2014):

$$\% \text{ Indel} = \left( 1 - \sqrt{1 - \left( \frac{(a + b)}{(a + b + c)} \right)} \right)$$

Donde  $a$  y  $b$  corresponden a la concentración relativa de las bandas cortadas y  $c$  equivale a la concentración relativa de la forma sin digerir del producto de PCR.

Locus	Región Amplificada	Cebador Directo	Cebador Reverso
EGFR	chr7:55,254,045-55,254,485	CCCCTCATGGCTTTCCCATC	ATTTGTTGGCAAACCTTGGGGA
EGLN3	chr14:34,407,413-34,408,207	AGATCCCGCATTTTCATGCAC	AAGCCTTTAGAGAGCCATAGGTG
PGK1	chrX:77,359,301-77,359,819	CCTCGAATTCCACGGGGTTG	GAGAGAGGTCCGGTGATTCCGG
EGFR + (Control positivo)	chr7:55,253,785-55,254,536	CAGGCCACTCAGGCCTATTC	CCTCACTATAGCTGCGCCTC

**Tabla 7. Cebadores utilizados para amplificación del ADN en el ensayo Surveyor.** Se indican los locus analizados, la región del ADN amplificada y los cebadores utilizados en la PCR.

#### iv) CRISPR: Edición del genoma

Se transfectaron, tal cual se comentó previamente, el plásmido que contiene la endonucleasa Cas9 y el ARN guía junto con el ADN donador. Hemos utilizado ADN donador de cadena simple de 200pb de longitud ubicando la modificación a introducir en el centro de la construcción (Integrated DNA Technologies). Este ADN donador, para el caso de EGFR introduce el SNP Chr7:55.254.186:A y en EGLN3 se muta el HRE descrito de EGLN3 chr14:34.407.696-34.407.700 modificando ACGTG a GCTAG. Para el locus de PGK1, el ADN donador tiene mutado los HRE1 y HRE2, y en su lugar se han incorporado los sitios de corte de BamHI y NcoI indicados en azul y verde, respectivamente en la [Tabla 8](#). La introducción de los sitios de corte para estas enzimas de restricción nos permite analizar la presencia de clones editados por recombinación homóloga realizando una digestión con enzimas de restricción.

Locus	ADN donador
EGFR	GCCTACCCCTGCTGACTAACTGTGGCCCCAGGGTGGTGACTCAGCCCTGCACCTCCTGATCCCGTCTGCACTGGGCCAGAGA GGATGACTTACCCAGCACATTCACATCACACAGCTTTGTGGATTCTAGGTCCAAGGACCAGAGATTTCAGTTATGTGAGTT ATTTTTTTATTTGTTCTTGCGTATTCCACAAAGGGT
EGLN3	ACATCGCCATATTAATACAGGGTGTGGGTTTGGGCGAGGCCGCCCTTCCCGTGGAGGACTGGCTCTAAGCCAGCAGGGTG ATGGCTGAGTGACAGGGCTGCTAGCGCTGCGTGAGGGTGGCAGCCAGCCACACCCGAGCACACCCGAGGCAGCCTGAG CTGGGAACGGACACTGGGAACCATCCCCCTCCCCATCTTACC
PGK1	CTTCGCCGTACCTTGTGGGCCCCCGGCGACGCTTCCTGCTCCGCCCTAAGTCGGGAAGGTTCTTGCGGTTTCGCGGCGT GCCGGGGATCCCAAACGGAAGCCCATGGCTCACTAGTACCTCGCAGACGGACAGCGCCAGGGAGCAATGGCAGCGCGC CGACCGCATGGGCTGTGGCCAATAGCGGTGAGCAGG

**Tabla 8. ADN donadores utilizados para edición del genoma mediante CRISPR/Cas9.** Se muestran la secuencias de ADN de simple cadena utilizado como ADN donador para favorecer la reparación del genoma vía recombinación homóloga. En rojo están señalados los nucleótidos modificados y esta subrayada la localización del motivo RCGTG en el genoma de referencia. En el donador de PGK1 se indica en azul y verde los sitios de corte para BamHI y NcoI, respectivamente que han sido insertados sustituyendo el HRE1 Y HRE2 del promotor de PGK1.

Para estimar el porcentaje de recombinación homóloga, solo en el caso de PGK1, en paralelo al ensayo Surveyor se amplificó la zona promotora con los cebadores indicados en la [Tabla 8](#), y el producto de PCR se digirió con BamHI o NcoI durante 1 hora a 37°C, tras lo cual se resolvió en un gel de agarosa. Nuevamente se utilizó la formula indicada en el trabajo de Pyzocha et al (Neena K. Pyzocha et al. 2014) para calcular, en este caso, el porcentaje de recombinación homóloga.

#### 4.11 Análisis de asociación entre SNPs identificados y los fenotipos de EPOC y SAOS.

Para realizar los estudios de correlación se construyeron tablas de contingencia para cada una de las variantes genéticas identificadas y los fenotipos de cada una de las enfermedades. Para estos análisis se han descartado aquellas variantes que: 1. Estaban presentes en <10% o >90% de los pacientes. 2. Variantes determinadas en menos de 50 muestras. 3. INDELs, es decir solo se trabajó

con variantes de un único nucleótido. Respecto de los fenotipos se han utilizado comorbilidades, fenotipos y escalas de pronóstico para EPOC, mientras que en SAOS se utilizaron comorbilidades (Tabla 9 y 10). En ambos casos los datos fueron proporcionados por los Doctores Francisco García Río del Hospital La Paz y Julio Ancochea del Hospital La Princesa de Madrid. En todos los casos se utilizaron datos anonimizados.

Para cada tabla de contingencia se utilizó el test de Fisher para determinar si existía asociación o no entre genotipos y fenotipos, obteniendo un p-valor para cada combinación de genotipo-fenotipo para cada paciente de EPOC y SAOS.

Los p-valores indicados no han sido corregidos con un test de comparaciones múltiples. Por ello, las asociaciones positivas se han interpretado con reserva esperando su validación en una muestra independiente de tamaño adecuado.

EPOC		Abrev.
FENOTIPO	Enfisema	Phe
	Bronquitis crónica	
	Agudizador con bronquitis crónica	
	Agudizador con enfisema	
	No agudizador con bronquitis crónica	
	No agudizador con enfisema	
	Mixto	
COMORBILIDADES	Cáncer	Can
	Cardiopatía Isquémica	CI
	Riesgo cardiovascular	RCV
	Hipertensión pulmonar	HTPCP
	Hipertensión arterial	HTA
ESCALAS DE GRAVEDAD/ PRONÓSTICO	Clasificación de gravedad Grado: I, II, III, IV	Grado
	Clasificación según la limitación del flujo aéreo. Global initiative for chronic Obstructive Lung Disease GOLD: A,B,C,D	GOLD
	Clasificación de Gravedad: 1 al 10	BODE
	Índice de comorbilidad <4, ≥ 4	COTE

**Tabla 9. Fenotipos de pacientes de EPOC.** Se muestran los fenotipos de los pacientes con EPOC utilizados para los estudios de asociación con las variantes genéticas identificadas.

Comorbilidades SAOS	Abrev.
Hipertensión arterial	HTA
Síndrome Metabólico	Smet
Insuficiencia cardíaca	ICC
Hipertensión pulmonar	HTP
Tromboembolismo pulmonar	TEP
Cáncer	CAN
Depresión/Ansiedad	D/A
Dislipidemia	DL
Diabetes mellitus	DM
Cardiopatía Isquémica	CI

**Tabla 10. Fenotipos de pacientes de SAOS.** Lista de comorbilidades de los pacientes con SAOS, utilizadas para los estudios de asociación con las variantes genéticas identificadas.

#### 4.12 Programas informáticos y análisis estadístico de los datos

Los programas informáticos utilizados en este trabajo, como Bowtie (Langmead et al. 2009) y SAMtools (Li et al. 2009; Li 2011) son de uso libre y están disponibles en repositorios públicos de software (Github y sourceforge):

<https://github.com/BenLangmead/bowtie>

<http://sourceforge.net/projects/samtools/files/>

Además, en el laboratorio hemos escrito programas originales usando como lenguaje de programación Python (<https://www.python.org/>). En concreto para la búsqueda de variantes en motivos RCGTG y la generación de tablas combinando las variantes de todos los pacientes.

El análisis estadístico y las representaciones gráficas fueron realizados en R (<http://www.r-project.org>), un programa de uso libre para el análisis estadístico y gráfico.

Para las comparaciones simultáneas de más de dos valores medios, se utilizó el análisis de la varianza (ANOVA) de una vía con una comparación múltiple post-test Tukey. Para el análisis de frecuencias observadas (variables cualitativas) en el estudio se asociación entre los genotipos y cada uno de los fenotipos para EPOC ó SAOS se utilizó un test de Fisher.





## 5 RESULTADOS



## 5.1 Identificación de regiones que contienen sitios de unión de HIF en el genoma

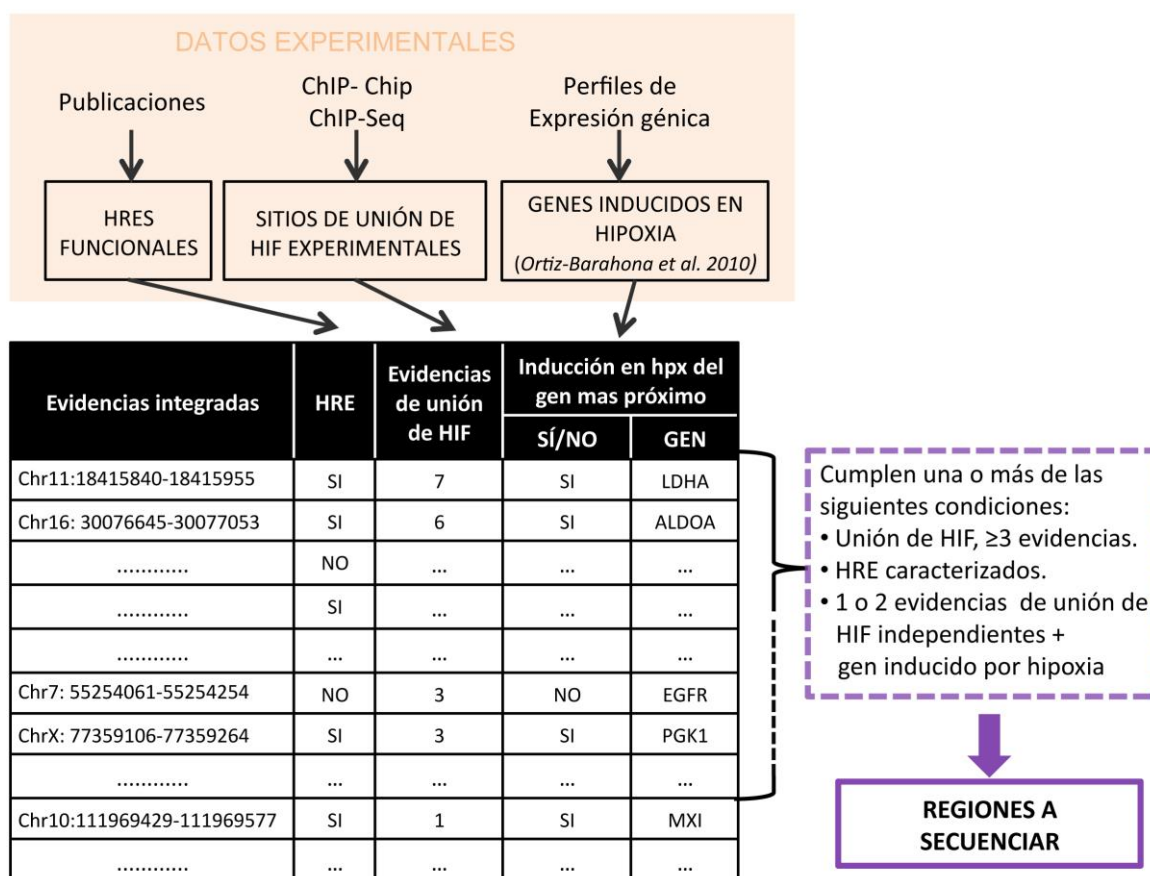
Los sitios de unión de HIF a nivel genómico han sido identificados por varios grupos de forma independiente (Mole et al. 2009; Schödel et al. 2012; Schodel et al. 2011; Villar et al. 2012; Xia & Kung 2009; Xia et al. 2009). Sin embargo, el análisis de estos resultados demuestra que el solapamiento entre las diferentes determinaciones experimentales es sorprendentemente pequeño (Villar et al. 2012). Esta concordancia limitada es probablemente consecuencia de las diferencias experimentales (tipos celulares, condiciones de hipoxia, etc.) y técnicas (diferentes plataformas y análisis) entre los diferentes estudios. Dado que el solapamiento entre los estudios publicados es mínimo y que no podemos abordar la secuenciación del genoma completo, nos propusimos diseñar una estrategia que nos condujera a determinar las regiones del genoma a las que HIF se une de forma general en la mayoría de los tipos celulares. Para ello utilizamos la información disponible en relación a los sitios de unión de HIF obtenida a partir de datos experimentales de tres tipos:

- a) HREs descritos en la bibliografía depositados en PubMed.
- b) Sitios de unión de HIF (HBSs, de HIF Binding Sites) determinados experimentalmente mediante inmunoprecipitación de cromatina combinada con secuenciación masiva de ADN (ChIP Sequencing) o inmunoprecipitación de cromatina combinada con *microarrays* de ADN (ChIP on chip)
- c) Perfiles de expresión génica de líneas celulares expuestas a hipoxia, de la base de datos pública Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>).

A partir de estos tres tipos de evidencias experimentales, se generó una tabla en la que para cada región potencial de unión de HIF se recogía el número de estudios en el que la región se encontraba unida a HIF, la presencia de HREs funcionales descritos y si alguno de los genes que flanquea la región se regula por hipoxia (Figura 12). De esta tabla se seleccionaron las regiones que cumplieran al menos uno de los siguientes criterios:

1. Unión de HIF determinada por ChIP-Seq o ChIP-chip en 3 o más estudios independientes.
2. Regiones con un HRE caracterizado en publicaciones en las que se estudia la regulación de genes individuales por hipoxia. Se seleccionaron solo HREs bien caracterizados, que incluían evidencias de ensayos reporteros, mutagénesis del HRE, ensayo de cambio en la movilidad electroforética y localizados en genes claramente implicados en la respuesta a hipoxia.
3. Regiones con una o dos evidencias experimentales de unión de HIF y asociadas a genes con clara inducción por hipoxia según el meta-análisis de perfiles de expresión génica realizado previamente en nuestro laboratorio (Ortiz-Barahona et al. 2010).

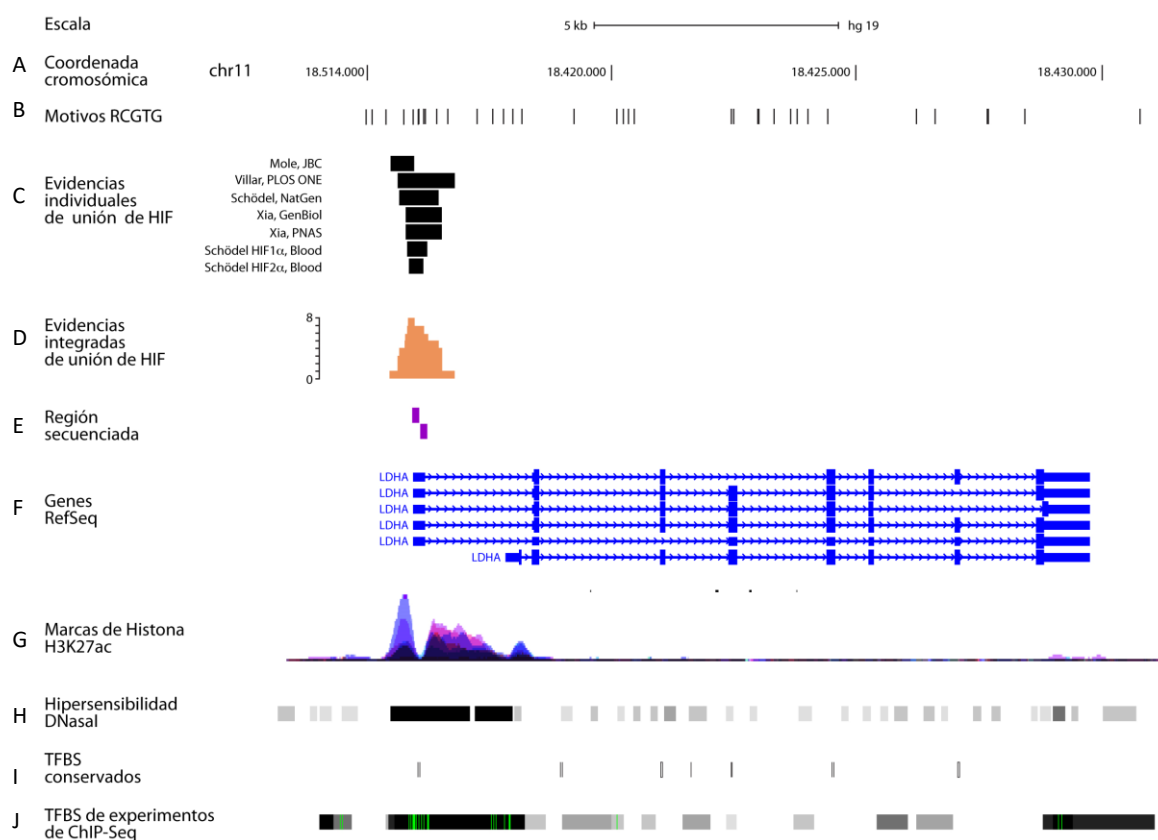
De este modo, obtuvimos un listado de 238 regiones reguladoras en el genoma humano que, con alta probabilidad, contienen HREs funcionales (Figura 12 y Anexo I).



**Figura 12. Esquema de la estrategia usada para definir las regiones a secuenciar que poseen sitios de unión de HIF.** Esta se basa en la integración de tres tipos de datos experimentales que ponen en evidencia la presencia de sitios de unión de HIF, permitiendo definir regiones reguladoras con sitios de unión de HIF.

Además, se investigó el potencial de estas regiones como reguladores de la expresión génica. Para ello, se examinaron los elementos funcionales indicativos de actividad transcripcional asociados a las regiones con potenciales HREs utilizando los datos generados por el proyecto ENCODE (Myers et al. 2011). En la Figura 13 se muestra como ejemplo el gen de lactato deshidrogenasa (*LDHA*) en el que se observa que la región seleccionada contiene un gran número de evidencias que sugieren una función reguladora de la transcripción. En concreto la región contiene marcas de histona H3K27 (H3K27Ac, indicadas en el carril G, que se asocian a regiones transcripcionalmente activas); áreas de hipersensibilidad a DNasI (carril H, que marcan cromatina abierta y por tanto accesible a la maquinaria de transcripción); regiones de unión de factores de transcripción a lo largo del genoma (TFBS conservados, carril I) y TFBS de experimentos ChIP-Seq, sitios de unión de factores de transcripción determinados por ChIP-Seq del proyecto ENCODE, en el carril J. El análisis de ésta

información nos permitió confirmar la funcionalidad de las regiones seleccionadas y delimitar con más precisión el fragmento a secuenciar.



**Figura 13. Selección de regiones reguladoras que poseen sitios de unión a HIF.** Como ejemplo se muestra la región reguladora del gen de la Lactato Deshidrogenasa (*LDHA*). Imagen obtenida del navegador genómico de la Universidad de California, Santa Cruz (UCSC). En la imagen se observa las evidencias de unión de HIF en el promotor del gen *LDHA*, así como las marcas de cromatina asociadas a regiones reguladoras transcripcionalmente activas. En concreto en esta imagen se observa: **A)** Coordenadas genómicas; **B)** Motivos RCGTG, indica todos los motivos RCGTG (R=A o G) presentes en la secuencia. **C)** Evidencias individuales de unión de HIF, en este caso se muestran las evidencias para el promotor *LDHA*, cada caja negra representa los datos de un experimento independiente de localización de HIF a nivel genómico. **D)** Evidencia integrada de unión de HIF, las cajas naranjas son la proyección de las múltiples evidencias indicadas en el apartado C a partir de la cual se determina la zona de máxima evidencia. **E)** Regiones secuenciadas, cada caja morada identifica uno de los amplicones de 250 pb secuenciados. **F)** Genes RefSeq, Transcritos de referencia. **G)** Marcas de histona H3K27ac, se asocia con la iniciación de la transcripción y cromatina de estructura abierta. **H)** Hipersensibilidad a DNase, indican zonas accesibles de la cromatina donde la cromatina pierde su estructura condensada y expone el ADN al corte por DNAsas. **I)** TFBS conservados, sitios de unión de factores de transcripción (TFBS) conservados en humanos, ratón y rata. **J)** TFBS de experimentos ChIP-Seq, sitios de unión de factores de transcripción determinados por ChIP-Seq del proyecto ENCODE.

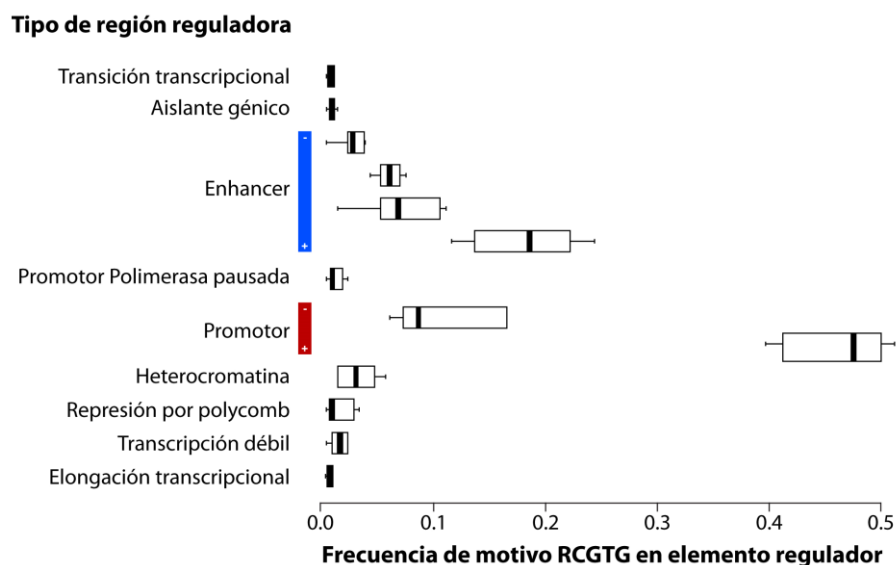
### 5.1.1 Caracterización funcional de los estados de la cromatina en regiones reguladoras con sitios de unión de HIF.

Con el fin de caracterizar el grupo de regiones seleccionadas globalmente, quisimos saber su distribución en relación a elementos funcionales del genoma. Para ello, utilizamos el carril

denominado “*estado de segmentación de la cromatina*” del proyecto ENCODE y el Broad Institute. Este carril utiliza un modelo de Markov oculto (HMM, de *hidden Markov model*) para clasificar (“segmentar”) cada región del genoma humano en “estados” específicos que caracterizan su función biológica (promotor, enhancer, etc.)(Ernst et al. 2011). Estos estados se determinan en base a datos experimentales de ChIP-Seq para nueve anticuerpos que se asocian a diversas características de la cromatina:

- Factor de transcripción CTCF. Es una proteína de unión al ADN que funciona como aislante génico bloqueando la actividad enhancer.
- Acetilación de la lisina 27 de la histona H3 (H3ac27), asociada al inicio de la transcripción y estructuras de cromatina abierta. Trimetilación de la lisina 27 de la histona H3 (H3K27me3), marca de promotores silenciados por proteínas Polycomb.
- Trimetilación de la lisina 36 de la histona H3 (H3K36me3), que marca regiones de elongación de la ARNPolIII, incluyendo transcritos codificantes y no codificantes.
- Metilación de la lisina 4 de la histona H3 (H3K4me), asociada a enhancers.
- Dimetilación de la lisina 4 de la histona H3 (H3K4me2), marca promotores y enhancers.
- Trimetilación de la lisina 4 de la histona H3 (H3K4me3), marca promotores activos y promotores.
- Acetilación de la lisina 9 de la histona H3 (H3K9ac), asociada al inicio de la transcripción y estructuras de cromatina abierta.
- Metilación de la lisina 20 de la histona H4 (H4K20me), asociada a regiones accesibles y activas.

En la [Figura 14](#) se muestra la distribución de las regiones seleccionadas para nuestro estudio en los diferentes estados funcionales de la cromatina. Más del 50% de las regiones corresponden a promotores y el segundo grupo funcional más representado son los enhancers donde se encuentra aproximadamente el 30% de las regiones que unen HIF. El 20 % de las regiones restantes se distribuyen entre las regiones funcionales del genoma que están menos asociadas a estados de cromatina activa como: transición transcripcional, aislantes génicos, promotores con polimerasa pausada, heterocromatina, regiones reprimidas por polycomb, transcripción débil y elongación transcripcional. Por lo tanto más del 80% de las regiones seleccionadas se localizan en elementos activadores de la transcripción, como corresponde a regiones con HREs funcionales.



**Figura 14. Distribución de motivos RCGTG presentes en las regiones secuenciadas según estados funcionales de la cromatina.** Se utilizó la predicción computacional de ENCODE llamada “estado de segmentación de la cromatina” para dividir las regiones a secuenciar en diferentes estados de la cromatina en relación a su función en la transcripción. Se observa que más del 50% de las regiones reguladoras corresponden a promotores, mientras que los enhancers constituyen el 30% de las regiones que unen HIF. El 20 % de las regiones restantes se asocian a regiones funcionales del genoma asociadas a estados de cromatina inactiva.

### 5.1.2 Estudio de ontología de genes asociados a las regiones reguladoras con sitios de unión de HIF.

Por otro lado, como un paso adicional de caracterización de las regiones seleccionadas, analizamos la función de los genes que regulan. Para ello, adscribimos cada uno de los 238 sitios de unión de HIF al gen más próximo (sitio de inicio de la transcripción más próximo) e identificamos las anotaciones funcionales enriquecidas en este grupo de genes utilizando el programa DAVID (DAVID, <http://david.abcc.ncifcrf.gov/>). Este programa extrae las funciones biológicas más representativas de un listado de genes dado, englobando en un mismo grupo funcional a aquellos que presenten funciones redundantes o similares (Huang et al. 2008). Así, clasificamos el listado de genes diana según el proceso biológico en el que estuvieran implicados (escogiendo la clasificación GOTERMBP\_ALL). El resultado de este análisis (Anexo II) puso de manifiesto que, los genes asociados a las regiones de unión de HIF están enriquecidos en funciones y rutas metabólicas reguladas por hipoxia (Figura 15) principalmente el metabolismo de los glúcidos y rutas de óxido/reducción.

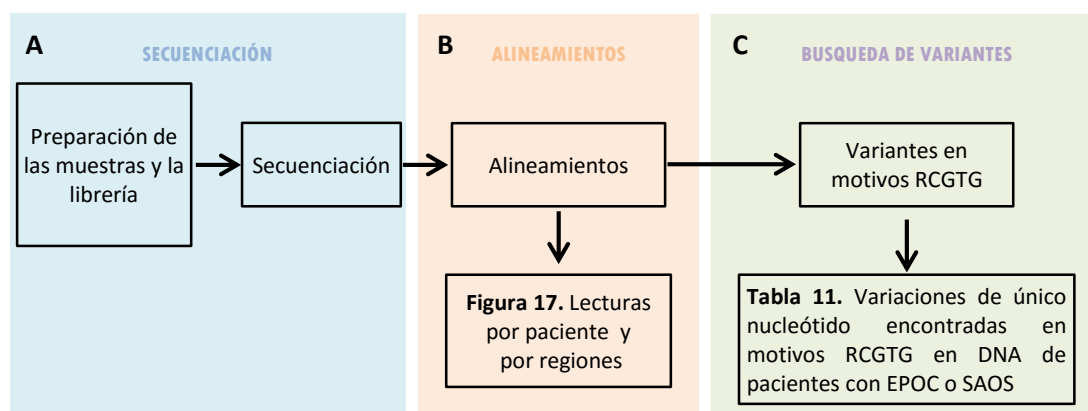
En conclusión, hemos identificado 238 regiones con alta probabilidad de unir HIF, que se localizan en regiones reguladoras de la transcripción (promotores y enhancers) y en la proximidad de genes implicados en la respuesta adaptativa a la hipoxia





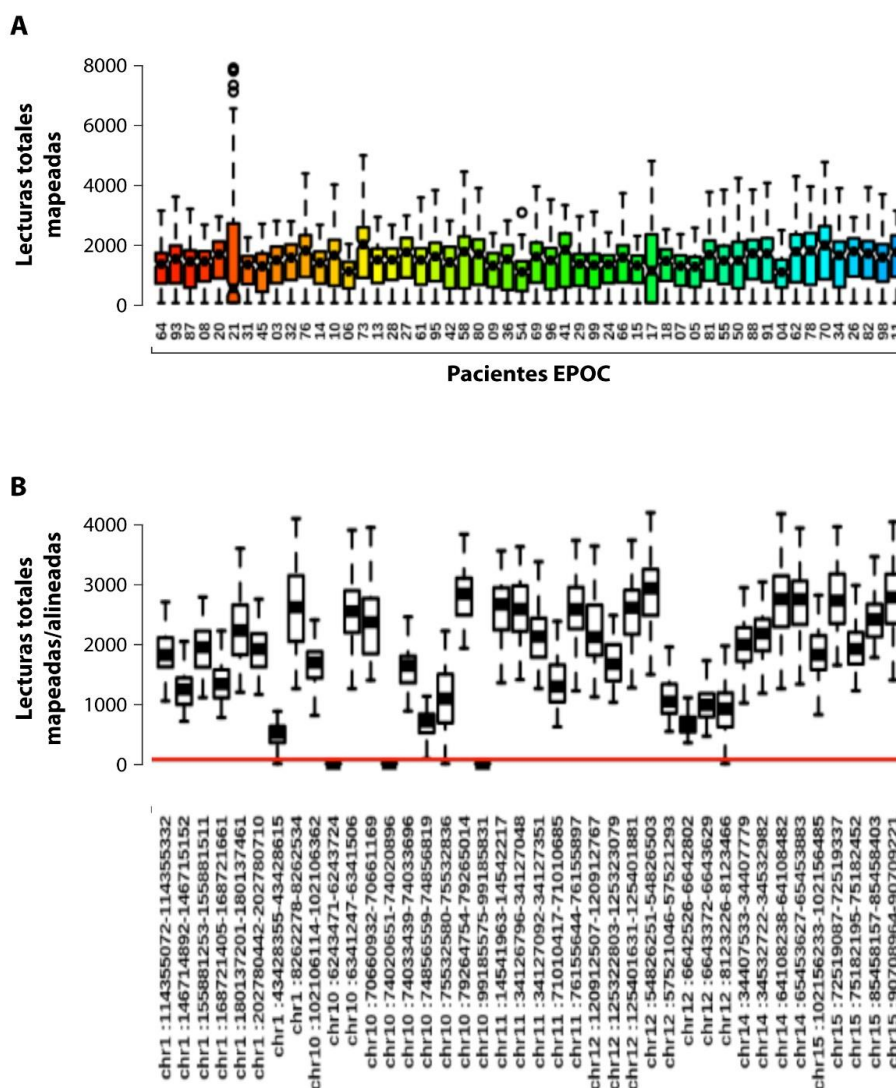
### 5.2.1 Identificación de variantes de un único nucleótido en sitios de unión de HIF por secuenciación masiva.

Un conjunto de 96 pacientes de EPOC y 144 pacientes de SAOS se seleccionaron de acuerdo a los criterios indicados en “Materiales y Métodos”. De cada paciente se obtuvo una muestra de ADN genómico extraído a partir de sangre periférica y se utilizó para generar una librería de las regiones reguladoras mediante *PCR Multiplex*. Las librerías se secuenciaron utilizando la plataforma MiSeq de Illumina, cuya tecnología se basa en la secuenciación por síntesis ([Figura 16A](#)).



**Figura 16. Esquema de secuenciación y búsqueda de variantes genéticas en motivos RCGTG.** Este proceso se divide en tres etapas: **A)** Secuenciación, que se inicia con la preparación de la muestra y generación de la librería y finaliza en la obtención de las secuencias; **B)** Alineamientos, a partir de los cuales se realizan estadísticas y controles de secuenciación; **C)** Búsqueda de variantes, identificación de variantes genéticas de un único nucleótido encontradas en motivos RCGTG en ADN de pacientes con EPOC o SAOS

Las lecturas resultantes de cada muestra se alinearon al ensamblado GRCh37/hg19 del genoma humano de referencia ([Figura 16B](#)). El número de lecturas obtenidas de cada una de las muestras fue similar entre individuos ([Figura 17A](#)). De la misma manera, al analizar el número de lecturas para cada una de las regiones secuenciadas en los pacientes de EPOC y SAOS, comprobamos que 196 de las 238 (82%) regiones secuenciadas presentaban una media aproximada de 1500 lecturas, más que suficientes para la determinación de variantes de forma fiable. En la [Figura 17B](#) se muestran 40 regiones representativas de las 238 secuenciadas en pacientes con EPOC.



**Figura 17. Lecturas por paciente y por región reguladora secuenciada.** A) Se muestran las lecturas mapeadas totales de 52 pacientes de EPOC representativas de los 98 pacientes secuenciados, donde se observa que el número de secuencias es similar entre pacientes. B) Se enseña el número de lecturas para 40 amplicones representativos de los 238 secuenciados. La mayoría de las regiones secuenciadas presentan un número de lecturas suficientes para la determinación de variantes genéticas de forma fiable. Sólo se observan unas pocas regiones que poseen un número de lecturas menor de 16, que se encuentran por debajo de la línea roja y resultan insuficientes para la determinación de variantes genéticas

A continuación, los alineamientos se rastrearon en búsqueda de variantes utilizando software estándar (Figura 16C). Del resultado de este análisis, solo nos quedamos con las variantes de un solo nucleótido y descartamos las inserciones/deleciones (INDELs) pues la detección de este tipo de variantes es poco fiable. A continuación, las variantes genéticas presentes en las regiones secuenciadas se clasificaron mediante un programa desarrollado en el laboratorio, en los siguientes grupos:

-Variantes que localizan en motivos RCGTG del genoma de referencia. Estas a su vez pueden ser de dos tipos:

- "MUT-HRE", variantes que destruyen motivos RCGTG existentes. Son cambios que afectan a la identidad de alguna de las cinco bases del HRE.
- "MUTREV", variantes que aun localizándose en un motivo RCGTG no lo destruyen. Por ejemplo el cambio de una adenina (A) por una guanina (G) en la primera posición del núcleo central del HRE. En este caso, probablemente, se modifica la afinidad de HIF por el motivo, ya que la frecuencia de una base en una posición puede reflejar afinidades diferentes, en este caso esperaríamos que si hay una A en lugar de G la afinidad sea mayor (Wasserman & Sandelin 2004).

-Variantes que no se localizan en motivos RCGTG del genoma de referencia. Una vez más existen dos posibilidades:

- "NEW-HRE", variantes que crean un nuevo motivo RCGTG en la región de interés.
- "NO HRE", variantes que no afectan a ningún motivo RCGTG ni crean uno nuevo.

Utilizando estos criterios hemos identificado un total de 72 variantes genéticas diferentes en regiones reguladoras que afectan motivos RCGTG en las muestras de pacientes de EPOC y SAOS secuenciadas, que se muestran en la [Tabla 11](#).

	Variante	Ref	Alt	GEN	SNP ID	EPOC	SAOS
MUT-HRE	chr1:116654495:A	ACGTG	ATGTG	<i>MABL123</i>	rs79725369		1
	chr1:168721546:A	ACGTG	ACATG	<i>DPT</i>	rs185593816	1	
	chr2:676899:T	CACGC	TACGC	<i>TMEM18</i>	No SNPs	1	
	chr2:74648825:T	CACGT	TACGT	<i>WDR54</i>	No SNPs		1
	chr2:74648828:A	CACGT	CACAT	<i>WDR54</i>	No SNPs	1	
	chr3:10157248:T	ACGTG	ACTTG	<i>BRK1</i>	No SNPs		1
	chr3:122102993:T	CACGC	TACGC	<i>FAM162A</i>	No SNPs		1
	chr3:183968282:T	ACGTG	ACTTG	<i>ECE2</i>	No SNPs		1
	chr3:4458679:A	ACGTG	AAGTG	<i>SUMF1</i>	rs114030216	1	
	chr5:121413872:T	ACGTG	ATGTG	<i>LOX</i>	No SNPs	1	
	chr6:126101997:A	GCGTG	GAGTG	<i>NCOA7</i>	rs114486929		1
	chr6:43736829:A	CACGC	CACGA	<i>VEGFA</i>	rs36208048	1	1
	chr7:100318291:T	CACGC	CATGC	<i>EPO</i>	rs191419026		1
	chr7:128095845:T	GCGTG	GTGTG	<i>HILPDA</i>	No SNPs	1	
	chr7:128095849:T	GCGTG	GTGTG	<i>HILPDA</i>	No SNPs		1
	chr7:154958601:G	CACGC	CACGG	<i>INSIG1</i>	rs546170887		2

	Variante	Ref	Alt	GEN	SNP ID	EPOC	SAOS
	chr7:154958704:T	CACGC	CATGC	<i>INSIG1</i>	rs1657290	24	63
	chr7:157197801:A	ACGTG	ACATG	<i>DNAJB6</i>	rs148894461		2
	chr7:157197920:A	GCGTG	GCATG	<i>DNAJB6</i>	rs1009329	8	11
	chr7:46018053:T	CACGC	CATGC	<i>IGFBP3</i>	rs555474543		1
	chr7:54826778:A	CACGC	CACAC	<i>SEC61G</i>	No SNPs		1
	chr7:54827021:G	CACGC	CGCGC	<i>SEC61G</i>	No SNPs		1
	chr7:55254186:A	CACGT	CACAT	<i>EGFR</i>	rs6593210	40	64
	chr7:74508875:C	ACGTG	ACGCG	<i>GTF2IRD2B</i>	rs587757601		1
	chr8:28223457:T	CACGT	CATGT	<i>ZNF395</i>	rs28492743		1
	chr10:102106247:T	GCGTG	GCTTG	<i>SCD</i>	rs115949610		1
	chr10:3089315:A	GCGTG	GCATG	<i>PFKP</i>	rs7078831	89	125
	chr10:3089424:A	ACGTG	ACGTA	<i>PFKP</i>	rs6602019	90	133
	chr10:75532772:T	ACGTG	ATGTG	<i>FUT11</i>	No SNPs		1
	chr10:93392984:T	ACGTG	ATGTG	<i>PPP1R3C</i>	No SNPs	1	
	chr11:10329736:C	GCGTG	GCGCG	<i>CAND1.11</i>	rs35389379		1
	chr11:71010448:G	CACGC	CAGGC	<i>SHANK2</i>	rs35853657	5	5
	chr11:71010489:C	ACGTG	ACCTG	<i>SHANK2</i>	rs181486845		3
	chr11:73882414:G	ACGTG	AGGTG	<i>PPME1</i>	rs72984898	2	5
	chr12:8123306:A	GCGTG	GCGTA	<i>SLC2A3</i>	rs7307261	84	124
	chr12:8123411:T	GCGTG	GCTTG	<i>SLC2A3</i>	rs368146797		1
	chr16:14458482:A	CACGC	CACAC	<i>PARN</i>	rs560640456		1
	chr17:18086364:T	CACGC	CATGC	<i>ALKBH5</i>	No SNPs		1
	chr18:13726682:C	GCGTG	GCGTC	<i>RNMT</i>	rs45561441	19	25
	chr18:77439276:T	GCGTG	GTGTG	<i>CTDP1</i>	No SNPs		1
	chr18:77439756:A	CACGT	CACAT	<i>CTDP1</i>	rs56246515	16	47
	chr19:16683332:G	CACGT	CGCGT	<i>SLC35E1</i>	No SNPs	1	
	chr19:16683441:C	ACGTG	ACGCG	<i>SLC35E1</i>	No SNPs	1	
	chr19:41903699:G	ACGTG	ACGGG	<i>BCKDHA</i>	rs45500792	26	30
	chr19:4968886:A	ACGTG	ACATG	<i>KDM4B</i>	No SNPs		2
	chr20:50366859:A	GCGTG	GCATG	<i>ATP9A</i>	rs118151281	9	10
	chr21:36260622:T	CACGC	CATGC	<i>RUNX1</i>	rs148801505	1	1
	chr22:30819999:T	CACGC	CATGC	<i>SEC14L2</i>	rs12165816		2
	chr22:35768219:G	CACGT	CGCGT	<i>HMOX1</i>	No SNPs	1	
	chrX:77359565:A	CACGT	CACAT	<i>PGK1</i>	No SNPs		1
	chrX:77359566:A	CACGT	CACGA	<i>PGK1</i>	rs150921338	1	4
MUTREV-HRE	chr5:131563631:G	CACGC	CACGC	<i>P4HA2</i>	No SNPs		1
	chr7:154958660:T	CACGC	CACGT	<i>INSIG1</i>	rs1730190	14	32
	chr16:30077265:G	CACGT	CGCGT	<i>ALDOA</i>	No SNPs	1	
	chr17:7343539:GC	CACGC	CACGC	<i>FGF11</i>	No SNPs	1	2
	chr19:16683438:G	ACGTG	GCGTG	<i>SLC35E1</i>	No SNPs	1	

	Variante	Ref	Alt	GEN	SNP ID	EPOC	SAOS
NEW-HRE	chr1:116654572:C	GTGTG	GCGTG	<i>MAB21L3</i>	No SNPs		1
	chr1:168721577:C	GTGTG	GCGTG	<i>DPT</i>	rs10918972	89	133
	chr14:64108378:G	CACTC	CACGC	<i>WDR89</i>	No SNPs	1	
	chr16:30077137:A	CGCGC	CACGC	<i>ALDOA</i>	No SNPs		1
	chr17:25621091:A	CGCGC	CACGC	<i>WSB1</i>	No SNPs	1	
	chr17:636309:A	CGCGC	CACGC	<i>FAM57A</i>	No SNPs		1
	chr17:636313:A	CGCGC	CACGC	<i>FAM57A</i>	No SNPs	1	
	chr17:80416090:G	ACGTC	ACGTG	<i>NARF</i>	rs28365943	62	93
	chr18:77439348:A	CGCGC	CACGC	<i>CTDP1</i>	No SNPs	1	
	chr18:77439608:A	CGCGT	CACGT	<i>CTDP1</i>	No SNPs	1	
	chr18:77439729:A	CGCGC	CACGC	<i>CTDP1</i>	No SNPs	1	
	chr19:48246933:A	CGCGC	CACGC	<i>EHD2</i>	No SNPs		1
	chr3:152878393:G	CCGTG	GCGTG	<i>RAP2B</i>	rs192987417		3
	chr7:105925641:G	GCGCTG	GCGTG	<i>NAMPT</i>	No SNPs	6	13
	chr7:128095816:G	ACGTG	GCGTG	<i>HILDPA</i>	No SNPs		3
	chr7:128095876:G	GCGTA	GCGTG	<i>HILDPA</i>	rs2732393	91	144

**Tabla 11. Variaciones de un único nucleótido identificadas en regiones reguladoras en motivos RCGTG en el genoma de pacientes con EPOC o SAOS.** Se muestran las variantes genéticas identificadas y clasificadas según que destruyan motivos RCGTG llamadas MUT-HRE; variantes en un motivo RCGTG que no destruyan el motivo, MUTREV-HRE o variantes que creen un nuevo motivo, NEW-HRE. Se indica para cada una de ellas su coordenada en el genoma, el motivo RCGTG de referencia (Ref.) y la secuencia con la variante alternativa (Alt), el gen más cercano, el SNP ID en los casos que exista, la frecuencia alélica menor (MAF), y el número de pacientes que tienen la variante génica en EPOC y SAOS. Las frecuencias alélicas se obtuvieron de la base de datos dbSNP build 142 excepto en \* donde se usó dbSNP build 141.

Al clasificar las variantes genéticas según se encuentren o no en motivos RCGTG encontramos la siguiente distribución:

-36 variantes en motivos RCGTG en pacientes de EPOC y 52 en pacientes de SAOS

-386 variantes fuera del motivo RCGTG en pacientes de EPOC y 444 en pacientes de SAOS (Tabla 12).

	EPOC	SAOS
Nº de pacientes	96	144
Variantes RCGTG		
MUT-HRE	25	41
NEW-HRE	11	11
Variantes NO RCGTG		
NO-HRE	382	441
MUTREV-HRE	4	3

**Tabla 12. Variantes genéticas identificadas en sitios de unión de HIF en pacientes de EPOC y SAOS.** En la tabla se presenta el número de pacientes de EPOC y SAOS secuenciados, así como el total de variantes genéticas para ambas enfermedades clasificadas según estén o no en motivos RCGTG y el tipo de cambio que produzcan.

En los pacientes con EPOC hallamos 25 variantes MUT-HRE, 11 NEW-HRE y 4 MUTREV-HRE. Cuando clasificamos las variantes genéticas para los pacientes de SAOS se hallaron 41 MUT-HRE, 11 NEW-HRE y 3 MUTREV-HRE (Tabla 12). El hecho de encontrar un número similar de variantes y que la distribución según la clasificación sea comparable entre EPOC y SAOS, a pesar de la diferencia en el número de pacientes para cada enfermedad, puede deberse principalmente a que las muestras provienen de la misma población (pacientes del Hospital de La Princesa y Hospital La Paz de Madrid). Además, encontramos 21 variantes genéticas en común en ambas enfermedades que corresponden en su mayoría a SNPs descritos, utilizando como referencia la base de datos *dbSNP* (Build142) del portal del NCBI ([www.ncbi.nlm.nih.gov/projects/SNP/](http://www.ncbi.nlm.nih.gov/projects/SNP/)). Por el contrario, aquellas variaciones que identificamos en una enfermedad y no en la otra son en general variantes de baja frecuencia.

A continuación comparamos las variantes identificadas por secuenciación masiva respecto de las esperadas en las regiones reguladoras secuenciadas de acuerdo a la base de datos *dbSNP* (Build142). En base a la información de *dbSNP*, esperabamos encontrar 285 SNPs comunes, que son aquellos cuya frecuencia alélica menor (MAF, del inglés Minor allele frequency) es  $\geq 1\%$ . Sin embargo, en la muestra de pacientes con EPOC y SAOS analizada hemos identificado 382 variantes genéticas y 441, respectivamente. Esta discrepancia se debe a que nuestros datos incluyen adicionalmente variaciones genéticas interindividuales de baja frecuencia en la población estudiada.

### 5.2.2 Validación de variantes genéticas identificadas en motivos RCGTG.

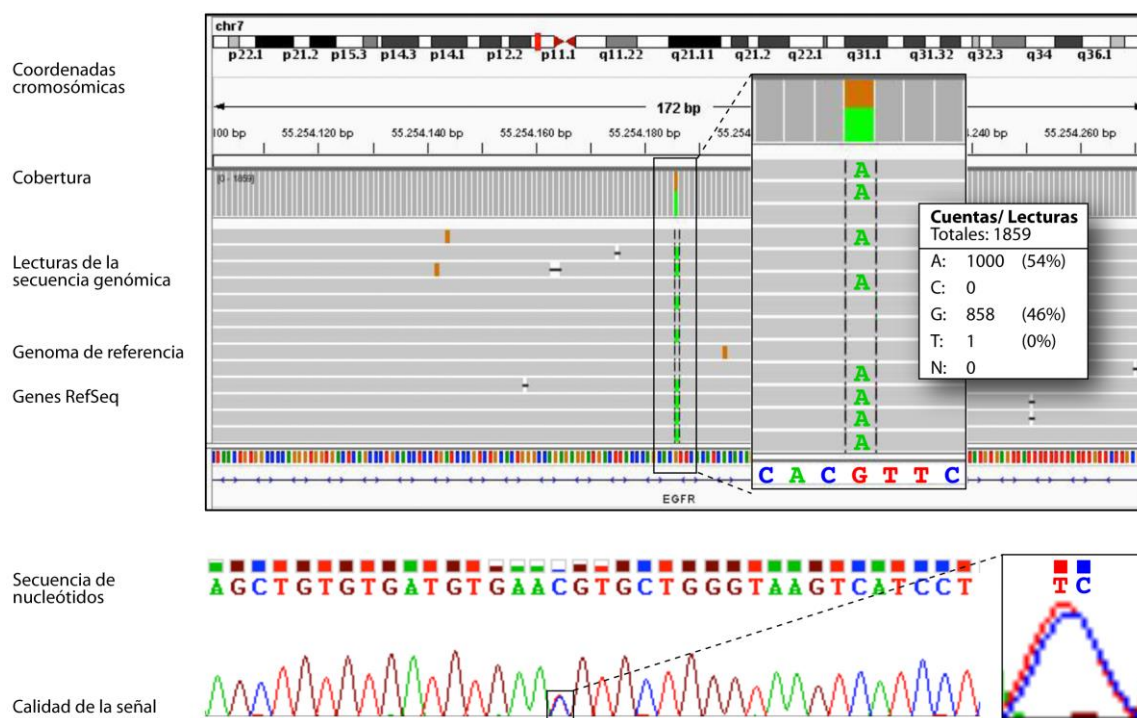
Con el propósito de validar los resultados obtenidos por NGS, se verificó la identidad de algunas de las variaciones genéticas mediante secuenciación de Sanger. Aunque la tasa de errores de la plataforma Illumina es menor del 0,4% y por ello esperamos un correcto genotipado de las muestras, consideramos importante realizar esta validación. En la Figura 17, se observa que se obtuvieron un número promedio de lecturas y calidad del alineamiento adecuados. Adicionalmente, al contrastar los resultados de NGS con los obtenidos por el método de Sanger, hemos observado que un 80% (16 de 20) de las secuencias validadas de EPOC fueron identificadas y clasificadas correctamente mediante NGS. En el 20% restante (4 de 20 secuencias) el error se debe tan solo a la determinación del número de alelos afectados por la variante. En la Tabla 13 se muestran dos ejemplos: i) Variante chr7:55254186:A que localiza en el gen de EGFR, como ejemplo de coincidencia y ii) Variante chr18:77439756:A que mapea en *CTDP1*, como ejemplo en el que no hay concordancia entre ambas técnicas de secuenciación en algunas muestras indicadas en negrita. Los casos en los que hemos detectado errores en el genotipado pueden deberse a regiones que no se amplifican bien en la PCR multiplex o a regiones en las que se favorece la amplificación de uno de los dos alelos. Por otro lado, los errores de genotipado coinciden con regiones que tienen un número de lecturas por debajo de la media como se observa en el caso de *CTDP1* en la Tabla 13.

Variante genética	GEN	Paciente	Sanger	Tabla	IGV % Variante Alt	Nº de lecturas totales
chr7:55254186: A	EGFR	EPOC02	0,5	0,5	54	1859
		EPOC08	1	1	99	1722
		EPOC09	1	1	100	1406
		EPOC43	1	1	100	1831
		EPOC51	1	1	100	1816
		EPOC04	0,5	0,5	49	1481
		EPOC01	0,5	0,5	47	1824
		EPOC02	0,5	0,5	54	1859
		EPOC21	0	0	0	142
		EPOC03	0	0	0	1862
		EPOC19	0	0	0	1936
		EPOC12	0	0	0	1666
chr18:77439756:A	CTDP1	EPOC05	0,5	0,5	71	182
		EPOC10	0,5	1	84	392
		EPOC17	1	0,5	59	102
		EPOC18	0,5	1	98	630
		EPOC24	0	0	0	458
		EPOC29	1	1	100	114
		EPOC32	1	1	91	268
		EPOC55	0,5	0,5	17	71
		EPOC56	0,5	0,5	57	106
		EPOC60	0,5	1	96	125
		EPOC62	0,5	1	99	361
		EPOC69	0,5	0,5	63	16
		EPOC76	0,5	0,5	18	232
		EPOC77	0,5	1	100	224
		EPOC84	1	0,5	13	61
		EPOC93	1	0,5	90	61
		EPOC104	0,5	1	100	64

**Tabla 13. Validación de variantes genéticas identificadas por NGS.** Para validar las variantes identificadas por NGS se secuenciaron por el método de Sanger 20 de las 72 variantes identificadas en EPOC. Se determinó que el 80% de las secuencias validadas se han genotipado correctamente. En la tabla se muestran una de las 16 variantes correctamente genotipadas como es el caso de chr7:55.254.186: A asociado a *EGFR*, y por otro lado se muestra la variante chr18:77.439.756:A asociada a *CTDP1* en la que hemos encontrado errores en el genotipado realizado a partir de los datos de NGS.

En la [Figura 18A](#), se muestra la región que incluye la variante genética chr7:55.254.186:A asociado a *EGFR* tal y como la representa el visor genómico *Integrative Genomics Viewer* (IGV). En esta imagen se observan las lecturas derivadas de NGS, procedentes de un paciente heterocigoto, alineadas al genoma de referencia. La [Figura 18 B](#), muestra el resultado de la secuenciación del mismo paciente por el método de Sanger. En esta imagen se visualiza el doble pico para la posición chr7:55.254.186 evidenciándose la condición heterocigoto.





**Figura 18. Visualización de la región próxima al SNP chr7:55.254.186:A en heterocigosis asociado a EGFR. A)** Imagen de visor genómico *Integrative Genomics Viewer* (IGV), en el que se observa la posición chr7:55.254.186:A en heterocigosis en la imagen ampliada y la región circundante. Se indica la localización de la región en coordenadas cromosómicas, el promedio de las lecturas de NGS en la cobertura, cada una de las lecturas independientes, el genoma de referencia y el gen RefSeq. En la tabla se indican el número de lecturas totales así como el número de lecturas asociadas a cada base nucleotídica. **B)** Captura de imagen de *Chromas 2.4*, en la que se muestra una región de un cromatograma obtenido por el método de Sanger. En la imagen ampliada se observa el doble pico para la posición chr7:55.254.186:A lo cual indica, puesto que el resto de la secuencia es simple, el genotipo heterocigoto para dicha posición.

Estos resultados confirman la fiabilidad de la secuenciación NGS para la identificación de variantes genéticas.

### 5.3 Caracterización funcional de SNPs en sitios de unión de HIF.

La resolución de los experimentos de ChIP-chip y ChIP-Seq está limitada a unas 300-1000pb; una longitud muy superior a los motivos de secuencia reconocidos por los factores de transcripción (5 bp en el caso de HIF). Por ello, la mayoría de las regiones seleccionadas para la secuenciación contienen más de un motivo RCGTG (Figura 13, carril B "Motivos RCGTG") y no se puede determinar cuál de los motivos media la unión de HIF a la región. Por ello, consideramos importante realizar ensayos que nos permitieran identificar los motivos RCGTG funcionales en cada región seleccionada y determinar los efectos de las variantes genéticas identificadas por NGS.



### 5.3.1 Estudio funcional del efecto de SNPs en motivos RCGTG mediante ensayos reporteros.

Para el análisis funcional de las variantes genéticas identificadas en los pacientes con EPOC y SAOS nos centramos en estudiar SNPs que destruyen motivos RCGTG (“MUT-HRE”) comunes a ambas enfermedades y que afectan a genes que se regulan por hipoxia. Para este último requisito, tomamos como referencia el meta-análisis de perfiles de expresión realizado previamente en nuestro grupo (Ortiz-Barahona et al. 2010) y la inducción mediada por qPCR-RT en un panel de líneas celulares que incluía: HeLa, HEK293T, IMR32 y MCF7 (datos no mostrados). Así, de los 15 SNPs MUT HRE comunes entre EPOC y SAOS, decidimos centrarnos en los nueve indicados en la [Tabla 14](#).

SNP	Cambio Ref/Alt	Gen	Meta-análisis Fold (p value)	HRE descrito	Zona reguladora
chr19:41.903.699:G rs45500792	ACGTG/ACGGG	BCKDHA	0,71( 0,0279)	NO	PROMOTOR
chr6:43.736.829:A rs36208048	CACGC/CACGA	VEGFA	3,43(2,78e-13)	SI, pero SNP no localiza en el descrito. chr6:43.844.950-43.844.954	
chrX:77.359.566:A rs150921338	CACGT/CACGA	PGK1	2,14 (5,38e-10)	3 HRES, SNP en HRE2 chrX:77.359.543-77.359.548 chrX:77.359.562-77.359.566 chrX:77.359.773-77.359.777	
chr12:8.123.306:A rs7307261	GCGTG/GCGTA	SCL2A3	0,14 (0,498) En micro array fold de 2,89	SI, pero SNP no localiza en el descrito. chr12: 8.123.422-8.123.426	ENHANCER
chr10:3.089.315:A rs7078831	GCGTG/GCATG	PFKP	2,29 (1,15e-05)	NO	
chr10:3.089.424:A rs6602019	ACGTG/ACGTA				
chr7:154.958.704:T rs1657290	CACGC/CATGC	INSIG1	1,95 (2,43e-06)	NO	
chr7:55.254.186:A rs6593210	CACGT/CACAT	EGFR	0,7 (0,13)	NO	
chr20:50.366.859:A rs118151281	GCGTG/GCATG	ATP9A	0,24 (0,446)	NO	

**Tabla 14. Perfiles de expresión de genes asociados a algunos de los SNPs identificados.** Se indican SNPs que son MUT-HRE, el tipo de cambio que producen, el gen asociado al SNP y la expresión génica en hipoxia de dichos genes a partir de datos del meta-análisis realizado previamente por nuestro grupo (Ortiz-Barahona et al. 2010). Por último se indica si el SNP afecta un HRE descrito, la posición del HRE descrito y el tipo de zona reguladora al que se asocia.

De ellos, la variante chrX:77.359.566:A, rs150921338, localiza en el promotor del gen de *PGK1* que codifica para la fosfoglicerato quinasa 1, una enzima involucrada en la glicolisis. *PGK1* esta regulado en hipoxia y contiene varios HREs funcionales en su promotor (Semenza et al. 1994; Semenza et al.

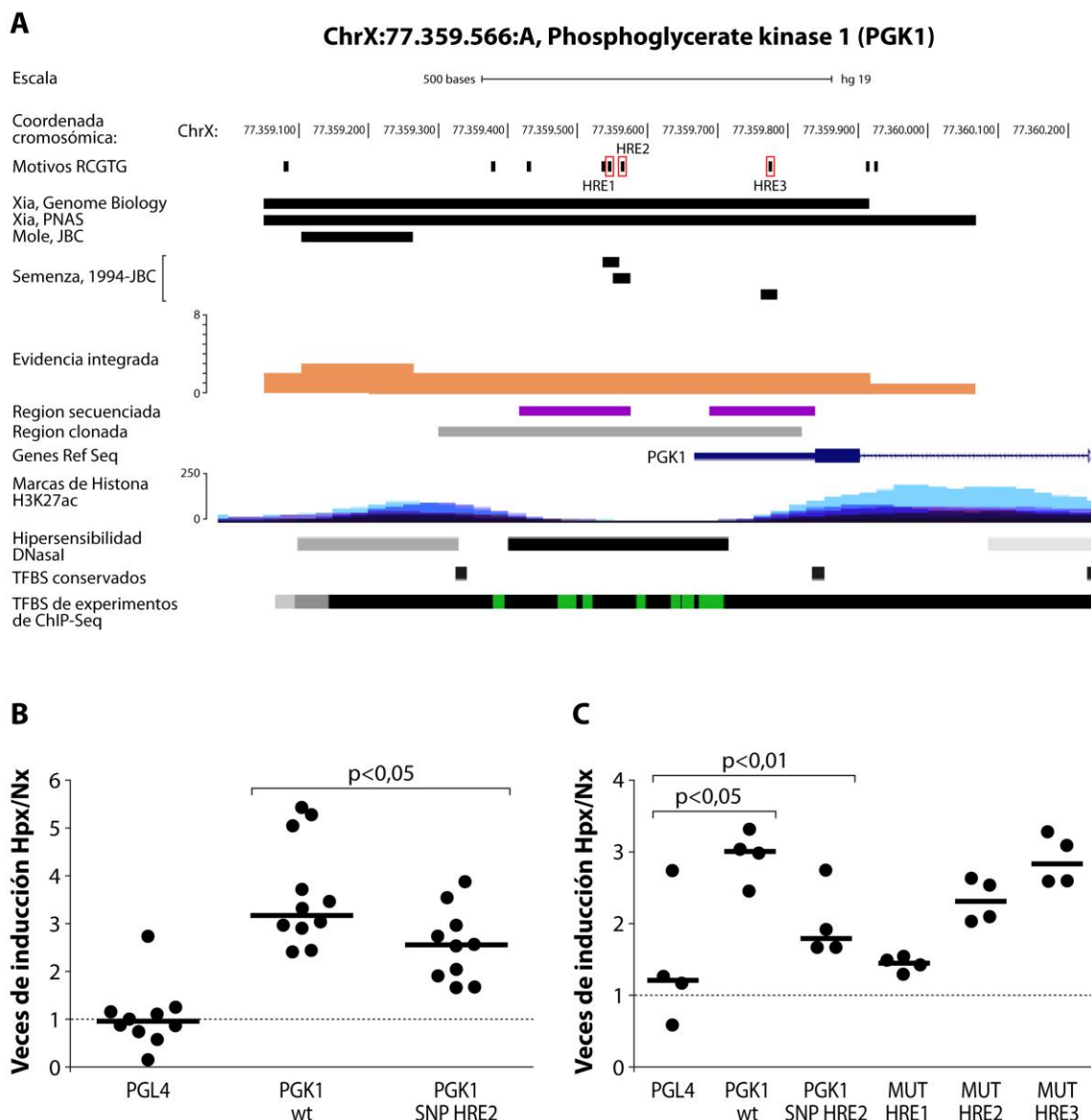
1996). Para el estudio de esta variante se clonó la región chrX:77.359.301-77.359.819, que contiene toda la región promotora de *PGK1*, en el vector reportero *pGL4.20*. Este mismo vector se utilizó para las demás variantes localizadas en promotores (tabla 14). En este y el resto de los casos, la región a clonar se seleccionó en base la información disponible del proyecto ENCODE sobre elementos funcionales del genoma. En concreto, escogimos regiones de cromatina abierta coincidentes con agrupamientos (“clusters”) de sitios de unión de factores de transcripción y con marcas de histonas asociadas a actividad transcripcional. La región reguladora de *PGK1* clonada contiene 3 motivos RCGTG. La variante asociada a esta región ChrX:77.359.566:A se encuentra en el HRE2 (Figura 19A) y destruye el motivo RCGTG cambiando de la secuencia original (CACGT en dirección reverse) la primera posición del núcleo de unión de HIF por una A (CACGA). Este SNP rs150921338 no está asociado a ninguna patología y su frecuencia en la población es de T: 98,553%; A: 1,447%.

Para generar las construcciones reporteras partimos de ADN genómico de un paciente heterocigoto, con el propósito de que las regiones reguladoras a evaluar por el ensayo de luciferasa solo varíen en el nucleótido de interés. De esta forma, para cada región reguladora se generaron dos construcciones, una con el nucleótido de referencia (wt) y otra con la variante alternativa (SNP). Se transfectaron cada una de estas variantes (wt o SNP), o el vector vacío *pGL4.20* como control, junto con un plásmido que codificaba para la enzima renilla en células HeLa y posteriormente fueron expuestas durante 16 horas a normoxia o hipoxia. Finalmente se midió la actividad renilla-luciferasa para cada transfección y los datos se procesaron corrigiendo con respecto a la renilla, como normalización de la eficiencia de transfección.

En la Figura 19B, se observa que la construcción *PGK1* wt se induce una media de 3,66 veces por hipoxia, mientras que la presencia del SNP chrX:77.359.566:A resulta en una inducción menor de 2,56 veces por hipoxia.

En el promotor de *PGK1* se han descrito tres HREs funcionales (Figura 19 A), que podrían explicar el efecto parcial del SNP sobre la inducción por hipoxia, por lo que analizamos mediante mutagénesis dirigida la contribución de cada uno de estos sitios. En la Figura 19C se muestra que el HRE1 (chrX:77.359.543-77.359.548) es el que más contribuye a la respuesta transcripcional pues su mutagénesis suprime por completo la inducción por hipoxia. Por el contrario, el HRE3 (chrX:77.359.773-77.359.777) no es funcional puesto que la destrucción del mismo no provoca diferencias en la inducción de la transcripción por hipoxia. Por otro lado, la mutagénesis del HRE2 (chrX:77.359.562-77.359.566) tiene un efecto parcial en la inducción de la construcción por hipoxia. Además, el efecto de esta mutación es comparable al observado en la construcción *PGK1* que presenta el SNP ChrX:77.359.566:A. Por lo tanto, la presencia de este cambio impide parcialmente la

regulación transcripcional de *PGK1* por parte de HIF, debido a que tanto HRE1 como HRE2 contribuyen a la inducción de *PGK1* por hipoxia.

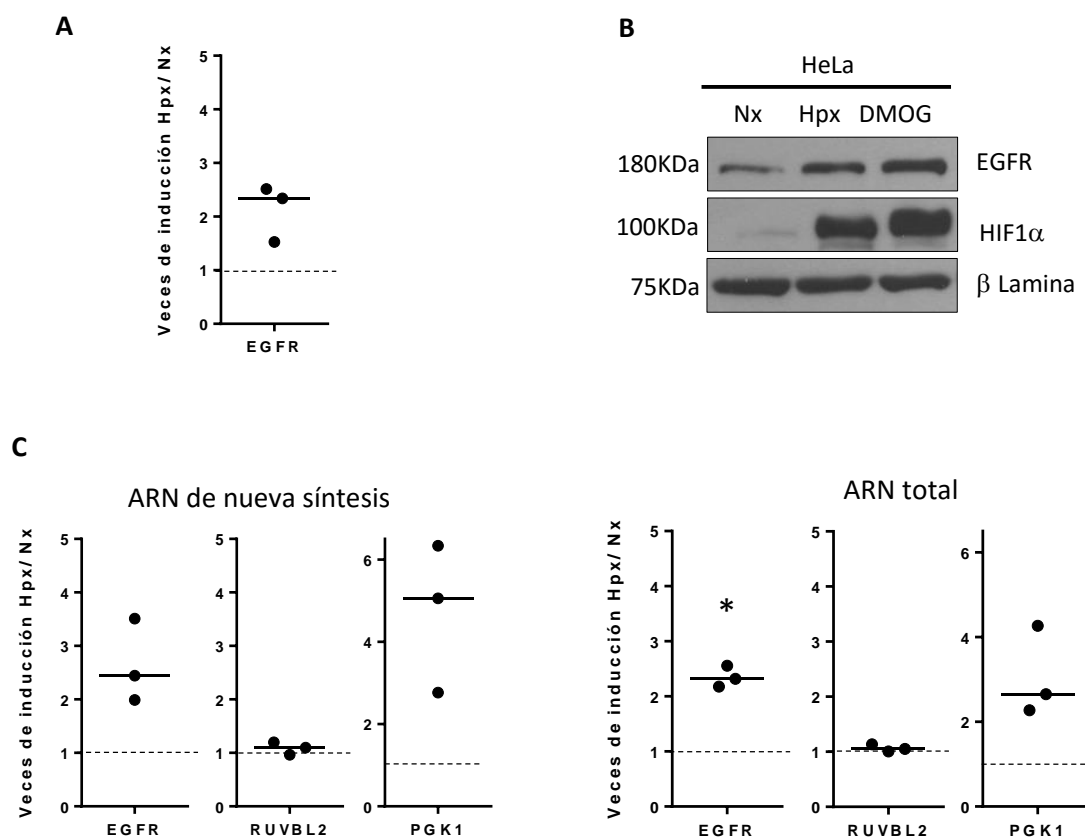


**Figura 19. Estudio funcional del efecto del SNP ChrX:77.359.566:A ubicado en el promotor de fosfoglicerato quinasa 1 (*PGK1*).** **A)** Imagen de UCSC de la región promotora de *PGK1* clonada para los ensayos reporteros donde se indican los tres HREs descritos. **B)** Ensayo de luciferasa para la región promotora de *PGK1*. Se muestra la versión wt para el HRE (*PGK1* wt) y la versión mutante que contiene el SNP ChrX:77.359.566:A (*PGK1* SNP HRE2) que provoca un efecto parcial de disminución de la inducción transcripcional. La diferencia entre los grupos es significativa (ANOVA,  $F_{2,26}=27,9$ ;  $p<0,01$ ) y un test de Tukey a posteriori demostró que la inducción media del genotipo wt y el que contiene el SNP son significativamente diferentes ( $p<0,05$ ). **C)** Ensayo de luciferasa en el que se ha realizado mutagénesis dirigida de los tres HREs descritos del promotor de *PGK1*. Se pone de manifiesto que el HRE1 y HRE2 son funcionales mientras que el HRE3 no posee actividad. La diferencia entre los grupos es significativa (ANOVA,  $F_{4,14}=8,8$ ;  $p<0,01$ ) y un test de Tukey a posteriori demostró que la inducción media del promotor conteniendo el SNP y la mutación del HRE1 son significativamente diferentes de las del wt ( $p<0,05$  y  $p<0,01$ ; respectivamente).

En el caso de las construcciones reporteras que contenían los promotores de *VEGFA* y *BCKDHA*, las inducciones por hipoxia de los wt fueron de pequeña magnitud, además las variantes identificadas en los HREs no modificaban la inducción (datos no mostrados), por lo que no se profundizó en su estudio.

Paralelamente, se estudió la variante chr7:55254186:A asociada al receptor del factor de crecimiento epidérmico (*EGFR*), este SNP rs6593210 no está asociado a ninguna patología y su frecuencia en la población es de G: 85%; A: 15%. La regulación de *EGFR* por hipoxia ha sido demostrada en varios trabajos previos. Sin embargo, se ha propuesto que esta regulación ocurre a nivel traduccional (Franovic et al. 2007). Por ello, antes de caracterizar el efecto del SNP en el motivo RCGTG en *EGFR*, estudiamos si *EGFR* se regulaba transcripcionalmente en hipoxia. En primer lugar, determinamos los niveles de ARNm mediante qPCR-RT en células expuestas a hipoxia y encontramos que, en nuestras condiciones, el ARNm de *EGFR* tiende a aumentar en hipoxia respecto de la normoxia (Figura 20 A). Además, la hipoxia aumenta los niveles de proteína EGFR que también son observados en el tratamiento con DMOG, que mimetiza las condiciones de hipoxia (Figura 20 B). Para determinar si la regulación de *EGFR* en hipoxia ocurre a nivel transcripcional realizamos un marcaje metabólico con 4-tiouridina (4SU) y analizamos las fracciones de ARN total y de ARN de nueva síntesis en condiciones de normoxia e hipoxia (Dölken et al. 2008). En la Figura 20 B, se observa que tanto el ARNm total como el ARNm marcado con 4SU (ARNm de nueva síntesis) se inducen por hipoxia, lo que sugiere que la acumulación del ARNm de *EGFR* se debe a un aumento de la transcripción del gen. En ambas fracciones observamos una tendencia de inducción del ARNm de EGFR en hipoxia, aunque el aumento es estadísticamente significativo solo para el ARNm total. Para este ensayo utilizamos *RUVBL2* y *PGK1* como control negativo y positivo, respectivamente, de inducción transcripcional génica en hipoxia.

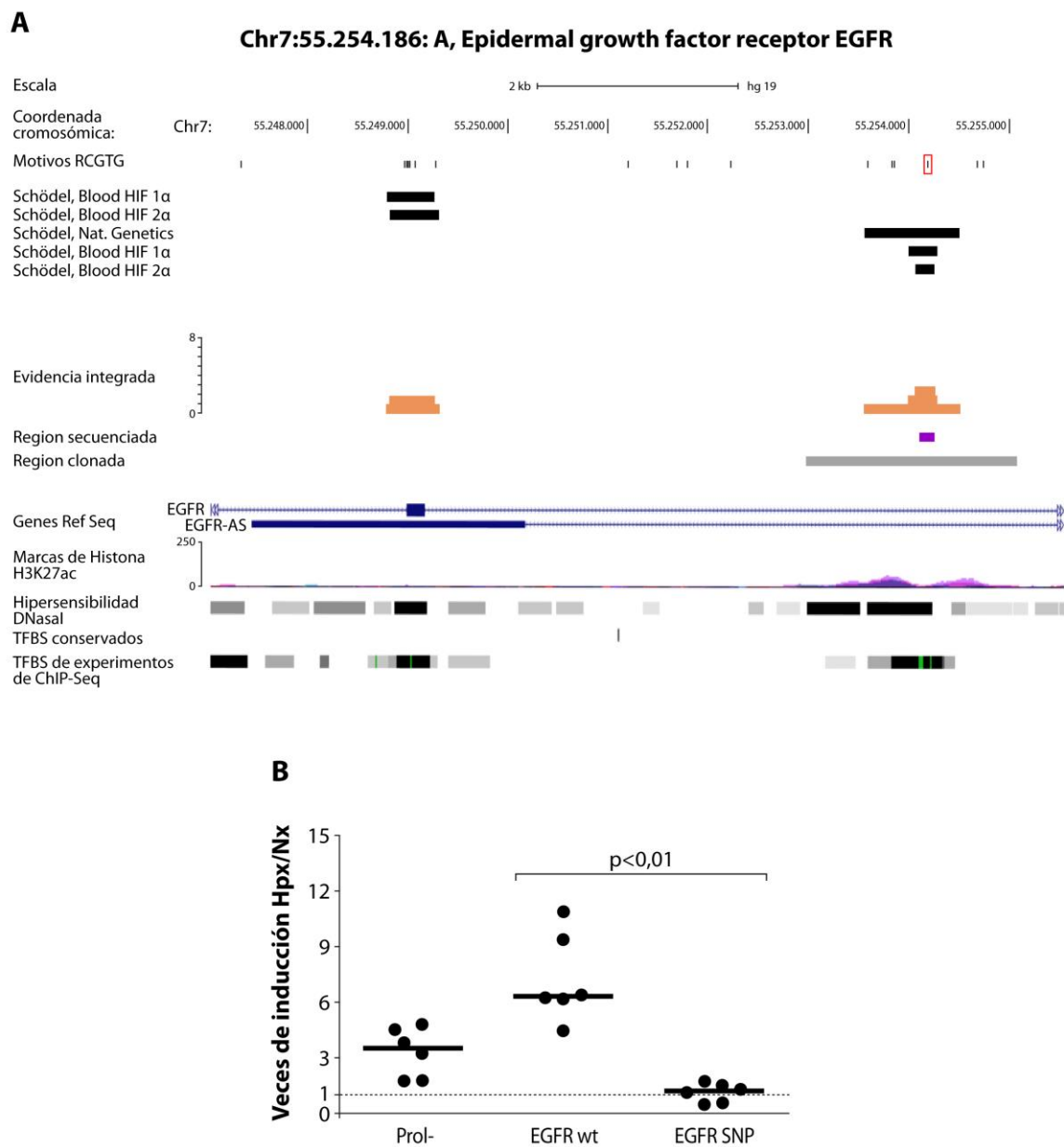
Una vez confirmamos que la transcripción de *EGFR* se induce por hipoxia, continuamos con el estudio de la región génica que contiene la variante chr7:55.254.186:A. Como se observa en la Figura 21 A, dicha variante se encuentra en un intrón del gen *EGFR* y por ello es posible que este RCGTG, de ser funcional, actúe como enhancer. Para el estudio de las variantes localizadas en enhancers (Tabla 14) se utilizó el vector *Prolactina* (Prol-) que contiene el gen de luciferasa bajo el control del promotor mínimo de prolactina de rata.



**Figura 20. Análisis de la regulación de la expresión génica de *EGFR* en hipoxia en células HeLa.** **A.** Análisis por qPCR-RT de los niveles de ARNm de *EGFR* en hipoxia versus normoxia. **B.** Análisis por *Western blot* de los niveles de EGFR, HIF1- $\alpha$  y  $\beta$ Lamina, en extractos totales de proteína de células HeLa expuestas durante 16 horas a normoxia, hipoxia o DMOG. El DMOG se utiliza como un mimético químico de la hipoxia pues inhibe la actividad de las PHDs. **C.** Análisis por qPCR-RT de los niveles de ARNm, tras ensayo de perfil de expresión génica de alta resolución que permite aislar ARN de nueva síntesis. Se observa el análisis por qPCR-RT de los niveles de ARNm de nueva síntesis y totales de *EGFR*, *RUVBL2* como un control negativo Y *PGK1* como control positivo de la regulación transcripcional en hipoxia. Las barras representan media  $\pm$  DE. La diferencia significativa entre grupos se indica con valores de probabilidad (p) del test de Student (\*,  $p < 0,05$ ).

Los datos de luciferasa demuestran que la región chr7:55.252.960-55.255.097 que contiene el motivo RCGTG en estudio, responde de forma robusta a hipoxia (Figura 21 B). Al evaluar el efecto del SNP rs6593210, EGFR Chr7:55.254.186:A se observa que la presencia del SNP elimina por completo la inducción génica en hipoxia respecto de la secuencia reguladora de referencia, *EGFR* wt (Figura 21 B). Así, la variación de un único nucleótido en el motivo RCGTG, que supone la destrucción del sitio de unión de HIF (CACGT a CACAT), suprime la inducción en hipoxia de *EGFR* en los seis experimentos independientes realizados.

Estos datos en su conjunto demostraron que la funcionalidad de un HRE puede verse afectada parcialmente o totalmente por la mutación de una única base en el motivo RCGTG.

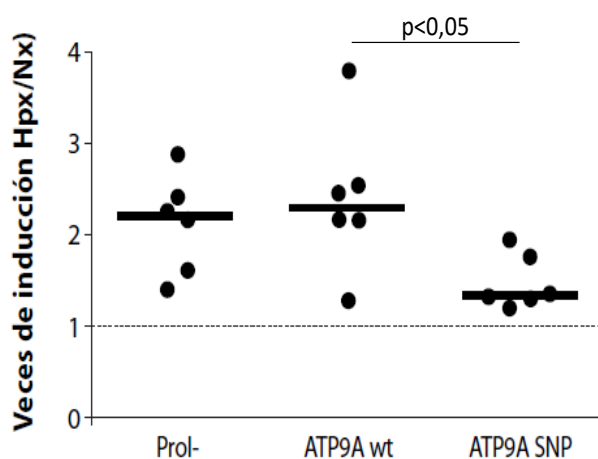


**Figura 21. Estudio funcional del efecto del SNP Chr7:55.254.186:A localizado en un intrón del gen del receptor del factor de crecimiento epidérmico (*EGFR*). A.** Imagen de UCSC de la región enhancer de *EGFR* clonada para los ensayos reporteros, donde se indica el HRE en el que se encuentra el SNP. **B.** Ensayo de luciferasa para la región enhancer de *EGFR*, en su versión wt para el HRE (*EGFR* wt) y la mutante (*EGFR* SNP) que contiene el SNP Chr7:55.254.186:A en la que se pierde por completo la inducción en hipoxia de la construcción reportera de *EGFR*. La diferencia entre los grupos es significativa (ANOVA,  $F_{2,14}=21,3$ ;  $p<0,01$ ) y un test de Tukey a posteriori demostró que la inducción media de la construcción reportera que contiene el SNP es significativamente diferente de la wt ( $p<0,01$ ).

En el caso del resto de variantes localizadas en enhancers (Tabla 14) encontramos que las regiones clonadas tenían una inducción por hipoxia de pequeña magnitud (unas 2 veces de inducción o menos) lo cual, unido al hecho de que el vector vacío (Pro-) presenta una inducción espuria de la misma magnitud, nos impidió obtener conclusiones de este bloque de experimentos. Como ejemplo, la Figura 22 muestra los resultados de la región chr20:50.366.359-50.367.325 correspondiente al primer intrón de *ATPase, class II, type 9A* (*ATP9A*) y en la que se localiza la variante chr20:50.366.859:A que corresponde al SNP rs118151281, que no se encuentra ligado a ninguna patología y su frecuencia en la población es de G: 98,6%; A: 1,4%. La hipoxia aumenta 2,4 veces la transcripción mediada por el enhancer con la secuencia wt de *ATP9A*; mientras que la que contiene la variación se induce 1,48 veces (Figura 22). Pese a que esta diferencia es estadísticamente significativa, no nos permite concluir que el SNP identificado en *ATP9A* afecta a la inducción por hipoxia, debido a que el vector vacío sufre una inducción aparente de 2,12 veces similar a la encontrada para la secuencia wt.

Así, aunque los resultados sugieren que, esta y otras de las regiones enhancer indicadas en la Tabla 14 se inducen por hipoxia y que la inducción se suprime por las variantes, los resultados no son concluyentes y sería necesario utilizar un vector reportero alternativo para analizar los SNPs presentes en regiones enhancer.

Considerando las limitaciones que hemos encontrado en los ensayos reporteros decidimos utilizar un abordaje alternativo basado en la edición del genoma mediante el sistema CRISPR/Cas9.



**Figura 22. Estudio funcional del efecto del SNP chr20:50.366.859:A ubicado en un intrón del gen que codifica para ATPasa, Clase II, Tipo 9A (*ATP9A*).** Ensayo de luciferasa para la región asociada a *ATP9A*, en su versión wt para el HRE (*ATP9A* wt) y la mutante (*ATP9A* SNP) que contiene el SNP chr20:50.366.859:A. Se observa un efecto parcial del mutante respecto del wt que no es concluyente porque el vector vacío se induce en una magnitud equivalente al wt en hipoxia. La diferencia entre los grupos es significativa (ANOVA,  $F_{2,14}=3,7$ ;  $p<0,05$ ) y un test de Tukey a posteriori demostró que la inducción media del promotor conteniendo el SNP es significativamente diferentes de la del wt ( $p<0,05$ ).

### 5.3.2 Análisis funcional del efecto de SNPs mediante edición del genoma en motivos RCGTG utilizando el sistema CRISPR/Cas9.

Con el propósito de editar el genoma para introducir las variantes genéticas identificadas y analizar sus efectos funcionales en respuesta a hipoxia se utilizó el sistema CRISPR/Cas9. Este sistema consiste en generar cortes de doble cadena en el ADN de manera dirigida y permite editar de forma precisa el genoma mediante la integración por recombinación homóloga de un ADN donador en el que se incorpora la variante genética de interés.

La importancia de utilizar el sistema CRISPR/Cas9 como un abordaje experimental alternativo a los ensayos reporteros, radica en que esta metodología nos permite editar el genoma de manera específica y precisa para analizar el efecto de cada SNPs en su contexto genómico.

Para realizar la edición del genoma utilizamos la línea celular HEK293T por ser una línea celular humana no transformada y por lo tanto con menos alteraciones cromosómicas que las líneas tumorales. Se han realizado las siguientes ediciones:

a) Introducción del cambio Chr7:55.254.186:A equivalente al SNP asociado a *EGFR*. Esta línea modificada se generó con el propósito de confirmar los resultados obtenidos en los ensayos reporteros.

b) Mutagénesis de los HRE1 y HRE2 del promotor de *PGK1*. Aunque el fin último es evaluar el efecto del SNP chrX:77.359.566:A, comenzamos editando el genoma para confirmar que estos elementos son necesarios para la inducción de *PGK1* en el contexto del genoma completo (Figura 19C). Para ello se modificaron ambos HREs mutando el motivo RCGTG y a la vez introduciendo sitios de corte para BamHI y NcoI en los sitios que corresponden a HRE1 y HRE2, respectivamente. La introducción de estos sitios de corte para enzimas de restricción, que no están presentes en esta región en el genoma de referencia de HEK293T, permitirá evaluar la edición del genoma por recombinación homóloga, así como facilitar el genotipado inicial de los clones. Los cambios que se introdujeron son: HRE1 chrX: 77.359.543-77.359.547 (ACGTG/ GGATC) y HRE2 chrX: 77.359.562-77.359.566 (CACGT/CATGG).

c) Por último, como control se editó el HRE funcional del enhancer de *EGLN3*. Este HRE en *EGLN3* fue descrito por nuestro grupo en un trabajo donde se comprobó que la construcción mutante del HRE no es capaz de inducir la expresión de luciferasa en condiciones de hipoxia (Pescador et al. 2005). El cambio que se intridujo fue: chr14:34.407.696-34.407.700 (ACGTG/GCTAG), que produce la alteración completa del motivo funcional RCGTG.



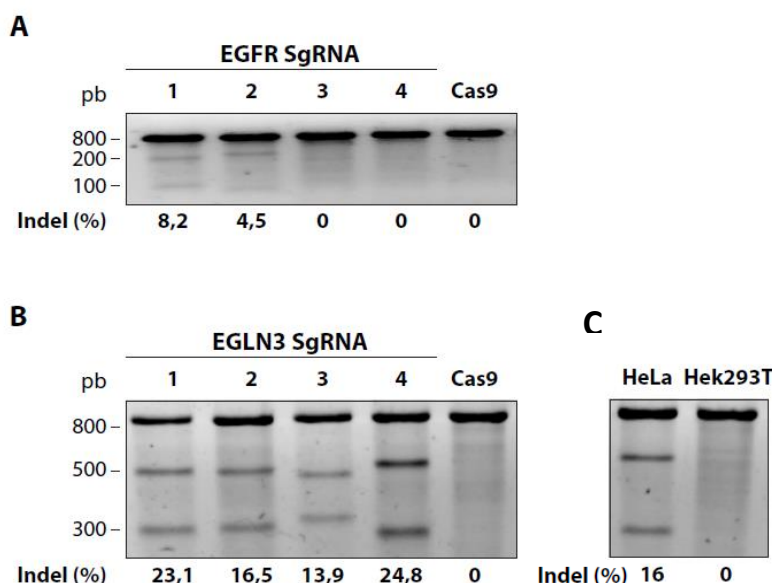
### Elección de ARN guía para edición de motivos RCGTG con el sistema CRISPR/Cas9.

Una vez seleccionadas las regiones a editar, el primer paso en el sistema de CRISPR/Cas9 es la elección del ARN guía, que es el elemento que da la especificidad a la edición. Puesto que se ha visto experimentalmente que para un mismo locus distintos ARN guía pueden dar lugar a eficiencias de corte significativamente diferentes (Mali et al. 2013), hemos diseñado 4 ARN guías para cada uno de los locus cuya diferencia radica en la posición que se encuentran respecto del sitio a modificar (Figura 23). Para evaluar la eficiencia de corte se realizaron transfecciones con Cas9 y cada uno de los ARN guías y pasadas 48 horas se realizó el ensayo Surveyor (ver “Materiales y Métodos”). Este ensayo, es una herramienta que evalúa si ha ocurrido o no edición del genoma en la población de células transfectadas y que permite determinar la eficiencia de los ARN guías.

EGFR	<p> <u>PAM</u>      <u>EGFR SgRNA 2</u>      <u>EGFR SgRNA 3</u>      <u>PAM</u>            CACTGGGCCAGAGAGGATGACTTACCCAG<b>CACGTT</b>CACATCACACAGCTTTGTGGATTCTAGGTCCAAGGACCA               PAM EGFR SgRNA 1      EGFR SgRNA 4 PAM         </p>
EGLN3	<p>              <u>EGLN3 SgRNA 2</u>      <u>PAM</u>      <u>PAM</u>      <u>EGLN3 SgRNA 3</u>            AGCAGGGTGATGGCTGAGTGACAGGGCT<b>ACGTG</b>CGCTGCGTGAGGGTGGCAGCCAGCCCACACCCGAGCACACCGCCAGGCA               EGLN3 SgRNA 4      PAM      EGLN3 SgRNA 1      PAM         </p>
PGK1	<p>              <u>PGK1 SgRNA 2</u>      <u>PAM</u>      <u>PAM</u>      <u>PGK1 SgRNA 4</u>            GGTTCCTTGCGGTTTCGCGGCGTGCCGG<b>ACGTG</b>ACAAACGGAAGCCG<b>CACGT</b>CTCACTAGTACCCTCGCAGACG               PGK1 SgRNA 1      PAM      PGK1 SgRNA 3         </p>

**Figura 23. Localización de ARN guías en los distintos locus a editar.** Se indica la localización de los ARN guías (Sg RNA) para cada uno de los locus. En negrita se señalan los sitios a modificar. En rojo se indica la secuencia PAM. Con una línea azul se indica la localización de los sgARNs en cada locus.

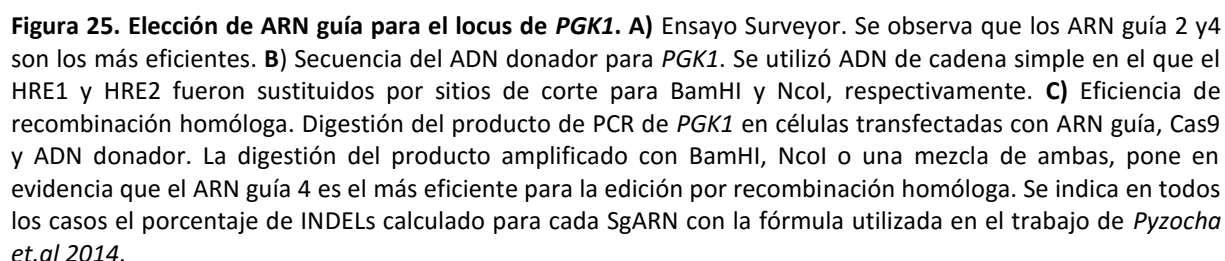
En la Figura 24, se muestran los resultados del ensayo Surveyor para los ARN-guías diseñados para editar el SNP de *EGFR* o eliminar el HRE funcional de *EGLN3*. Para el locus de *EGFR* el ARN guía 1 es el más eficiente (Figura 24 A), que genera un porcentaje de corte del 8,2%. El carril correspondiente a Cas9 es un control negativo para la edición del genoma, en el que se ha transfectado el vector PX459 (Addgene) en el que no se ha clonado ARN guía. Por tanto, teniendo en cuenta que la mejor eficiencia de corte de la enzima Surveyor implica un mayor porcentaje de células editadas por el sistema CRISPR/Cas9 hemos seleccionado el ARN guía 1 para *EGFR*.



**Figura 24. Ensayo Surveyor para los locus de *EGFR* y *EGLN3*.** **A)** Ensayo Surveyor para los diferentes guías utilizados en la edición del SNP de *EGFR* (indicados como carriles 1, 2, 3 y 4), en el que se observa que el ARN guía 1 es el más eficiente. El carril indicado como Cas9 corresponde a células transfectadas solo con Cas9 y sin ARN guía (Plásmido PX459 vacío). **B)** Ensayo Surveyor para los diferentes guías utilizados en la edición del HRE de *EGLN3* (indicados como carriles 1, 2, 3 y 4), en el que el ARN guía 4 es el más eficiente. El carril indicado como Cas9 corresponde a células transfectadas solo con Cas9 y sin ARN guía (Plásmido PX459 vacío). **C)** Ensayo Surveyor de la región chr7:55.253.785-55.254.536 en el locus de *EGFR* utilizada como control del ensayo, para la cual HeLa es heterocigoto para el SNP chr7:55.254.036 en mientras que HEK293T es homocigoto. Se indica en todos los casos el porcentaje de INDELS calculado para cada SgARN con la fórmula utilizada en el trabajo de Pyzocha et.al 2014.

Cuando realizamos el ensayo Surveyor transfectando las células HEK293T con los ARN-guías para *EGLN3*, hemos logrado la mejor eficiencia de corte y por tanto de edición del locus de *EGLN3* para el ARN guía 4 (Figura 24 B), para este caso el porcentaje de corte calculado es de 24,8%, por tanto fue seleccionado para editar el HRE funcional de *EGLN3*.

A su vez utilizamos como control de este ensayo Surveyor el locus chr7:55.253.785-55.254.536, cercano al SNP estudiado en *EGFR*, para el cual en la posición chr7:55.254.036 la línea celular HeLa es heterocigoto mientras que para HEK293T es homocigoto, de modo que tenemos un control positivo y negativo, respectivamente (Figura 24 C).



77

donde se concluye que el ARN guía 4 es el único para el cual se ven bandas como resultado de la digestión con las enzimas de restricción. Además hemos calculado, con la misma fórmula de Pyzocha (Pyzocha Neena et al. 2014) el porcentaje de recombinación homóloga, que es de 2,1% cuando se digiere con ambas enzimas de restricción.

#### **Edición de motivos RCGTG con el sistema CRISPR/Cas9.**

Como se indicó previamente, nuestro interés en la herramienta CRISPR/Cas9, es lograr editar el genoma de HEK293T de manera precisa para introducir los SNPs en posibles HREs funcionales identificados en el genoma de pacientes con EPOC o SAOS. Por ello, buscamos que la reparación del ADN una vez que corta la enzima Cas9 se produzca por recombinación homóloga. Para introducir los SNPs de interés en el genoma de HEK293T, a la mezcla de transfección de Cas9 y ARN guía se adicionó un ADN donador que incluye el SNP o mutación de interés para cada locus. Para este abordaje utilizamos ADN donador de simple cadena (cuyos brazos de homología son de 100 bases nucleotídicas a cada lado de la región a modificar). Para la edición del SNP identificado en *EGFR* en pacientes de EPOC y SAOS el donador solo varía del genoma de referencia en una base en la posición chr7:55.254.186 donde se ha sustituido una G por una A. Mientras que para la eliminación del HRE de *EGLN3* el ADN donador presenta un cambio en tres bases nucleotídicas en el motivo RCGTG cuyas coordenadas son chr14:34.407.696-34.407.700; de manera que el HRE de referencia ACGTG se modifica a la secuencia GCTAG, destruyendo el HRE funcional en las células que hayan sido editadas. Para el locus de *PGK1* el donador se realizó de modo similar, tal como se comentó previamente, sustituyendo los HRE 1 y HRE2 por los sitios de corte de restricción para las enzimas BamHI y NcoI (Tabla 8).

Así, una vez seleccionados los ARN guías y ADN donador para realizar la edición del genoma en los HREs funcionales en *EGFR*, *EGLN3* y *PGK1*, se transfectaron HEK293T con el ARN guía y ADN donador para cada locus. Pasadas las 48 horas se aislaron clones por dilución límite, que una vez amplificados fueron genotipados por secuenciación mediante el método de Sanger. En la Tabla 15 se indica el número total de clones aislados para cada genotipo, así como el número de clones que presentan INDELS, y aquellos que están editados por recombinación homóloga en un solo alelo (wt/SNP), o en ambos alelos (SNP/SNP). A partir de estos datos se pone de manifiesto que la edición de clones mediante CRISPR/Cas9 no ha resultado muy eficiente en nuestras manos, en particular el porcentaje de clones editados por recombinación homóloga es menor al 1% en todos los casos. Incluso no hemos sido capaces de aislar ni un solo clon editado para el locus de *PGK1*.

CLONES						
	NO Editados	Editados			Perdidos	Nº Total aislados
	wt	wt/INDEL	wt/SNP	SNP/SNP		
<i>EGFR</i>	123	3	3	1	6	136
<i>EGLN3</i>	73	9	2			84
<i>PGK1</i>	136	3*	-	-		139

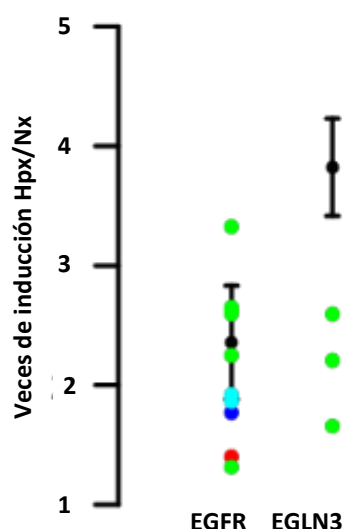
**Tabla 15. Número de clones aislados para la edición de los HREs de locus *EGFR*, *EGLN3* o *PGK1* con el sistema CRISPR/Cas9.** Se indica el número total de clones aislados y el desglose de estos según se encuentren editados o no. En el caso de los editados se clasificaron según el genotipo heterocigotos (wt/ SNP) y homocigotos (SNP/SNP). Para *PGK1* el número de wt/INDEL esta subestimado pues en este caso se genotipo mediante el corte con enzimas de restricción, poniendo en evidencia la edición vía HDR y no se cuantifican los INDELs con precisión.

A continuación se estudió el efecto del SNP chr7:55.254.186:A de *EGFR* y la delección del HRE funcional en *EGLN3* en la regulación de la inducción de la transcripción génica en hipoxia. Para ello, las líneas celulares editadas en los locus de *EGLN3* y *EGFR* se cultivaron durante 16 horas en normoxia o hipoxia y posteriormente se determinó la expresión de los genes de interés mediante qPCR-RT. Para el locus de *PGK1* no fue posible realizar este experimento porque no se obtuvieron clones editados como resultado de la baja eficiencia en la edición por CRISPR/Cas9.

La [Figura 26](#) muestra el efecto de la edición SNP en el intrón de *EGFR* y la delección del HRE funcional en *EGLN3* en la regulación de la inducción de la transcripción génica en hipoxia. Se muestra la inducción media de *EGLN3* por hipoxia de 15 clones en los que el HRE no está editado (wt/wt, negro). En la edición del HRE de *EGLN3*, los clones editados fueron todos heterocigotos (wt/INDEL, verdes), en estos clones la inducción por hipoxia del ARNm de *EGLN3* se reduce significativamente; ya que la inducción media es inferior al intervalo de confianza de los clones wt. Estos resultados confirman que el HRE editado en *EGLN3* es funcional y esencial para su inducción transcripcional y son la prueba de concepto que evidencia la validez de esta técnica para el estudio funcional de elementos de respuesta a hipoxia.

De forma análoga a la edición de *EGLN3* en la [Figura 26](#) se muestra la inducción media de *EGFR* por hipoxia de 8 clones en los que el SNP no está editado (wt/wt, negro). En la edición de este SNP se aislaron clones homocigotos (SNP/SNP (azul oscuro), INDEL/INDEL (rojo)) y heterocigotos (SNP/INDEL, azul claro; wt/SNP y wt/INDEL (verde)). Se observa que en los clones editados en homocigosis se reduce significativamente la inducción de *EGFR* en hipoxia, pues la inducción media

de todos los clones analizados es inferior al intervalo de confianza de los clones wt, mientras que en los clones heterocigotos la inducción de *EGFR* en hipoxia no es concluyente.

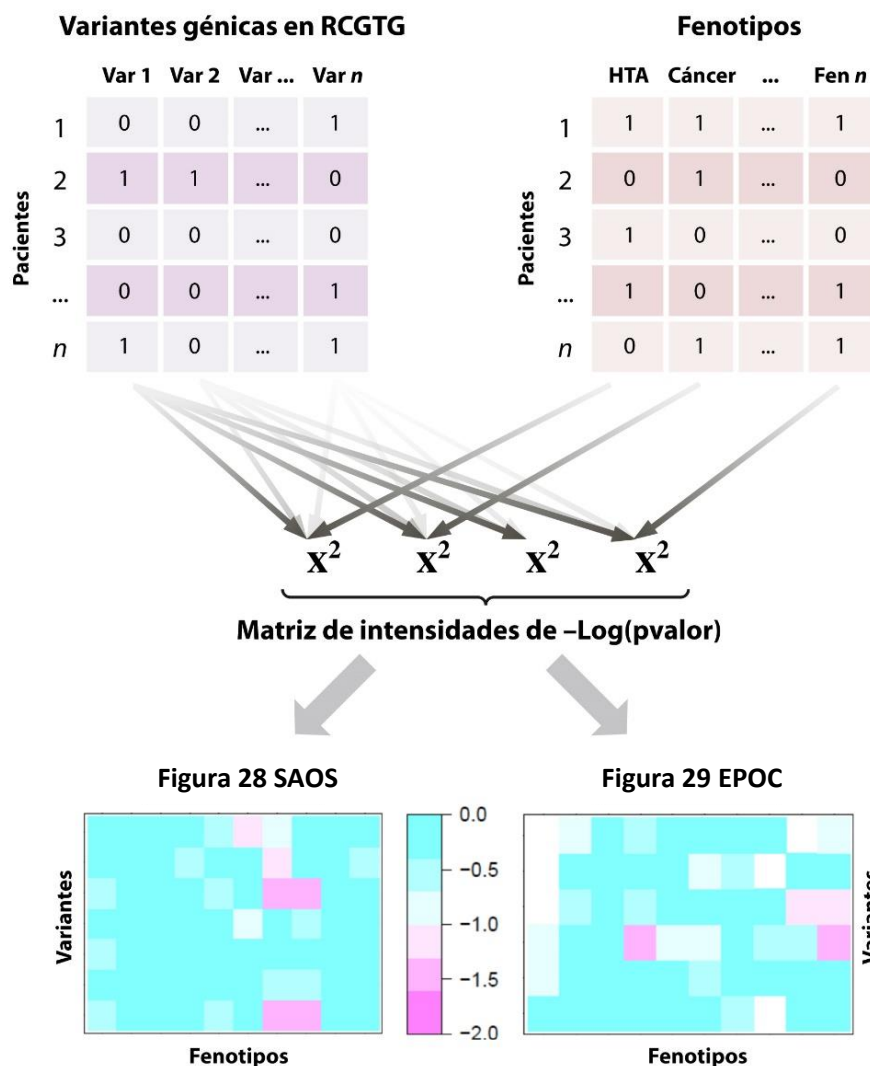


**Figura 26.** EL SNP Chr7:55.254.186:A altera la inducción transcripcional de *EGFR* por hipoxia en células HEK293T editadas por el sistema CRISPR/Cas9. La gráfica representa la expresión de *EGFR* y *EGLN3* en líneas celulares wt ó editadas con el sistema CRISPR/Cas9 mediante qPCR-RT. Cada punto de la gráfica muestra las veces de inducción de *EGFR* y *EGLN3* en hipoxia sobre los niveles en normoxia para los diferentes genotipos, verde: wt/SNP o wt/INDEL; azul claro: SNP/INDEL; azul oscuro: SNP/SNP; rojo: INDEL/INDEL. El valor mostrado en cada punto experimental es la media de 3-5 determinaciones independientes. El símbolo negro indica la media de inducción en los clones de genotipo wt (n=8 en el caso de *EGFR* y n=15 en *EGLN3*) y las barras de error representan el intervalo de confianza del 95% de la media.

En conclusión, estos resultados demuestran que los clones que poseen el genoma editado en homocigosis para el SNP chr7:55.254.186:A, tienen una menor inducción de *EGFR* en hipoxia. Por tanto, la modificación de un único nucleótido en el motivo RCGTG del HRE funcional de *EGFR* modifica la regulación de la expresión de este gen en condiciones de hipoxia.

#### 5.4 Correlación entre las variantes genéticas en sitios de unión de HIF y los fenotipos de la EPOC y el SAOS.

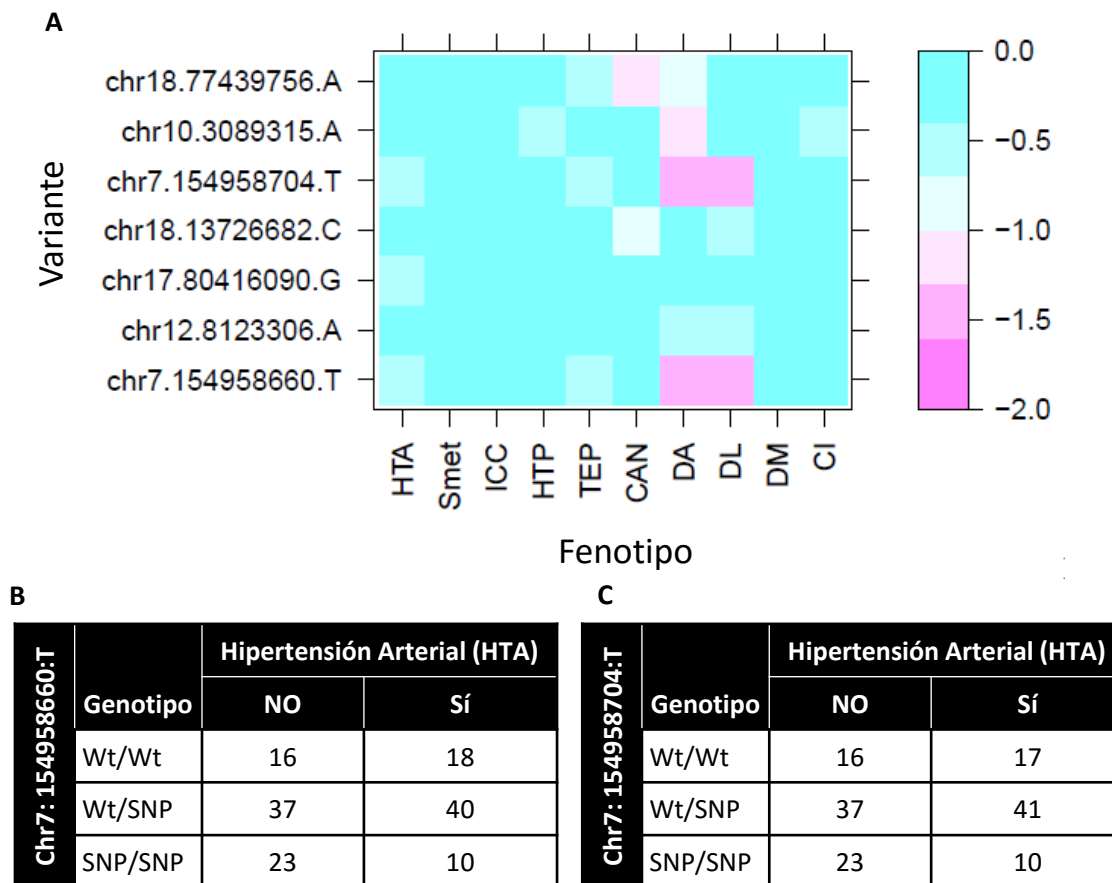
Una vez identificadas las variantes genéticas presentes en motivos RCGTG en pacientes con EPOC y SAOS, se realizó un estudio de correlación genotipo-fenotipo, con el fin de estudiar la existencia de asociaciones entre las variantes identificadas y los fenotipos de ambas enfermedades. Para ello, se seleccionaron las variantes que no tuvieran una distribución muy sesgada en la población estudiada. Así, seleccionamos variantes genéticas que estuviesen en al menos el 10% y en no más del 90% en las muestras de EPOC o SAOS. Además, eliminamos aquellas que no se hubieran secuenciado en al menos 50 pacientes. Seleccionamos un total de 6 variantes genéticas para el análisis de correlación en pacientes de EPOC y 7 en el caso de pacientes de SAOS.



**Figura 27. Esquema del análisis realizado para el estudio de correlación genotipo-fenotipo.** Se realizó el test de Fisher para cada una de las combinaciones de variantes genéticas identificadas y fenotipos o comorbilidades, para SAOS y EPOC. Se obtuvo un p-valor para cada una de las combinaciones y ese resultado se presenta en una matriz de intensidades para cada enfermedad, en la que se representa el negativo del logaritmo de los p-valores para las asociaciones entre las diferentes variantes genéticas en motivos RCGTG (genotipo) y los diferentes fenotipos.

Hemos trabajado con fenotipos, comorbilidades y escalas de pronóstico para EPOC y comorbilidades para SAOS proporcionadas por los colaboradores clínicos del Hospital La Paz y el Hospital La Princesa. Los estudios de asociación se realizaron utilizando el test de Fisher para cada una de las combinaciones de variantes y fenotipos posibles. Los resultados se muestran como una matriz de intensidades que representa el negativo del logaritmo de los p-valores resultantes del test de Fisher para cada combinación de fenotipo-genotipo (variantes genéticas en motivos RCGTG) (Figuras 27).

En el análisis de correlación genotipo-fenotipo en los pacientes de SAOS, no se han observado asociaciones significativas (Figura 28A). Sin embargo, se evidencia una tendencia de asociación entre el SNP chr7: 154.958.660:T y el SNP chr7: 154.958.704:T con la hipertensión arterial (HTA) (pvalor de 0,08515 y 0,08613 según el test de Fisher, respectivamente).



**Figura 28.** Análisis genotipo-fenotipo de los pacientes con SAOS. **A)** Matriz de intensidad del negativo del logaritmo de los p-valores. Se muestra los resultados para cada comparación, siendo los valores menores a -1,3 significativos por corresponder a un p-valor menor a 0,05. No se han encontrado asociaciones significativas para SAOS, pero se evidencia una tendencia de asociación entre el SNP Chr7: 154.958.660:T y Chr7:154.958.704:T con la hipertensión arterial (HTA). Los valores inferiores a -1,3 para cáncer (CAN), depresión/ansiedad (DA) y dislipidemia (DL) no corresponden a asociaciones fiables. **B)** Tabla de contingencia para la variante Chr7: 154.958.660:T y HTA, en la que se indica la distribución de pacientes. Se observa la tendencia de que los pacientes de SAOS homocigotos para este SNP tendrían una menor probabilidad de desarrollar hipertensión arterial que los wt. **C)** Tabla de contingencia para la variante Chr7: 154.958.704:T y HTA, en la que se indica la distribución de pacientes. Se observa la tendencia de que los pacientes de SAOS homocigotos para este SNP tendrían menor probabilidad de desarrollar hipertensión arterial que los wt.

HTA: Hipertensión arterial; Smet: Síndrome metabólico; ICC: Insuficiencia cardiaca; HTP: Hipertensión pulmonar; TEP: Tromboembolismo pulmonar; CAN: Cáncer; DA: Depresión/ansiedad; DL: Dislipidemia; DM: Diabetes mellitus; CI: Cardiopatía Isquémica.

En la tabla de contingencia para el SNP Chr7:154.958.660:T, rs1730190, y HTA en los pacientes de SAOS (Figura 28B), se observa una tendencia a la correlación entre la presencia del SNP rs1730190 en homocigosis y una menor probabilidad de padecer HTA (OR:0,386 ; IC95%: 0,142-1,053). Para el SNP Chr7:154.958.704:T, rs1657290, y HTA la tabla de contingencia es muy similar a la anterior (Figura



28C) por lo que también hay una menor probabilidad de padecer HTA cuando este SNP está en homocigosis (OR:0,409 ; IC95%: 0,149- 1,115). Aunque esta asociación no es estadísticamente significativa en la muestra de 144 pacientes de SAOS analizada, se observa dicha tendencia en las tablas de contingencia; pero sería necesario ampliar la muestra de pacientes de SAOS para consolidar esta correlación.

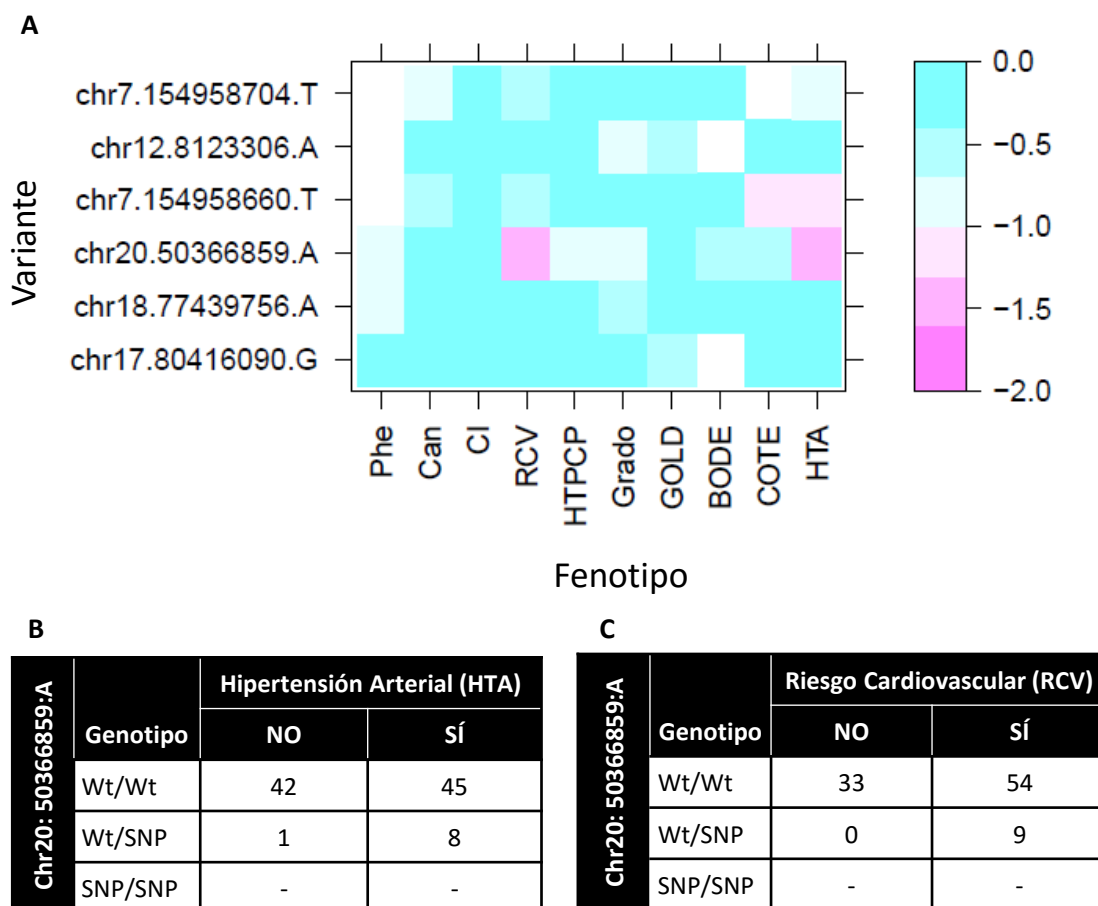
El SNP Chr7:154.958.660:T consiste en un cambio de C a T y el motivo RCGTG no se destruye (CACGC/CACGT), pues se trata de un MUTREV HRE en el que la primera posición del núcleo central de unión de HIF pasa de una G a una A, lo cual podría modificar la afinidad de la unión de HIF por el ADN. El SNP Chr7:154.958.704:T es un cambio de C a T y el motivo HRE se destruye. Por su localización genómica este SNP podría actuar como un enhancer, ya que la distancia entre el potencial HRE y el gen más próximo es de 79.050 pares de bases (*HTR5A*). Por ello, no nos es posible inferir sobre que gen podría estar actuando este motivo RCGTG en caso de ser in HRE funcional.

Los datos de correlación con cáncer (CAN), depresión/ansiedad (DA) y dislipidemia (DL) (Figura 28A) no corresponden a asociaciones fiables, ya que no se cuenta con un número de pacientes suficiente con estas comorbilidades.

Al realizar la correlación genotipo-fenotipo para la muestra de pacientes de EPOC, hemos encontrado dos asociaciones para la variante genética chr20:50.366.859:A, con hipertensión arterial (HTA) y la segunda con riesgo cardiovascular (RCV), siendo los p-valores 0,0392 y 0,0254, respectivamente (Figura 29A). La tabla de contingencia para la asociación de chr20:50.366.859:A y HTA se muestra en la Figura 29B, en la que se observa que la presencia de dicha variante en heterocigosis se asocia a una mayor probabilidad de padecer HTA (OR:7,467 ; IC95%: 0,896-62,265). La asociación encontrada entre la variante chr20:50.366.859:A y RCV se muestra en la Figura 29C, donde se observa que la presencia del SNP se asocia con una mayor probabilidad de padecer RCV (OR: infinito, pues hay una de las categorías que es cero).

La variante genética chr20:50.366.859:A para la que se encontró asociación a comorbilidades en EPOC corresponde a la variante asociada a *ATP9A*, previamente estudiada en los ensayos reporteros (Figura 22). Por tanto, hemos identificado una variante genética que se asocia a dos comorbilidades en pacientes de EPOC, que interrumpe un HRE funcional y podría afectar a la inducción de la transcripción en hipoxia. Al analizar la inducción por hipoxia de *ATP9A*, así como de los genes próximos a éste, no hemos hallado diferencias respecto de la normoxia (datos no mostrados); por lo que es posible que la regulación de *ATP9A* por hipoxia se dé solo en algunos tipos celulares o bien que este HRE controle la inducción de otros genes que no hemos analizado.

En conjunto estos resultados ponen de manifiesto la existencia de SNPs en regiones reguladoras, más concretamente en motivos de unión de HIF, que pueden asociarse a fenotipos de EPOC o SAOS y por tanto podrían explicar las variaciones en el desarrollo de estas enfermedades.



**Figura 29. Análisis genotipo-fenotipo de los pacientes con EPOC. A)** Matriz de intensidad del negativo del logaritmo de los p-valores. Se muestra los resultados para cada comparación, siendo los valores menores a -1,3 significativos por corresponder a un p-valor menor a 0,05. Se ha encontrado asociación para la variante Chr20: 50.366.859:A con hipertensión arterial (HTA) y con riesgo cardiovascular (RCV). La asociación para HTA tiene un p-valor de 0,0392 mientras que el p-valor para RCV es 0,0254. **B)** Tabla de contingencia para la variante Chr20: 50.366.859:A y HTA, en la que se indica la distribución de los pacientes de EPOC. Se observa una asociación entre la variante Chr20: 50.366.859:A e HTA para la que los individuos heterocigotos para el SNP tienen mayor probabilidad de desarrollar HTA en los pacientes de EPOC. **C)** Tabla de contingencia para la variante Chr20: 50.366.859:A Y RCV, donde se indica la distribución de los pacientes de EPOC. Se observa una asociación entre la variante Chr20: 50.366.859:A y una mayor probabilidad de padecer RCV en los individuos heterocigotos para el SNP.

Phe: Enfisema, Bronquitis crónica, Agudizador con bronquitis crónica, Agudizador con enfisema, No agudizador con bronquitis crónica, No agudizador con enfisema, Mixto; CAN: Cáncer; CI: Cardiopatía Isquémica; RCV: Riesgo cardiovascular; HTPCP: Hipertensión pulmonar; Escalas de gravedad/pronóstico: Grado, GOLD, BODE, COTE; HTA: Hipertensión arterial.





## 6 DISCUSIÓN



Los estudios de perfiles de expresión génica junto con los GWAS han demostrado que los polimorfismos tienen un papel crucial en la expresión génica diferencial en humanos (Schadt et al. 2003; Morley et al. 2004) y que sus efectos en *cis* o *trans* pueden estar implicados en rutas biológicas que están afectadas en enfermedades (Moffatt et al. 2007; Hunt et al. 2008).

Por otro lado, resultados obtenidos previamente en el laboratorio demostraron que la presencia de variantes en un HRE funcional pueden alterar la regulación génica inducida por hipoxia (Ortiz-Barahona et al. 2010), poniendo de manifiesto la importancia de la integridad de la secuencia consenso R(A/C)CGTG en la regulación transcripcional mediada por HIF.

En enfermedades respiratorias obstructivas como la EPOC y el SAOS, donde la hipoxia es una característica central, HIF tiene un papel esencial en la regulación génica (Yoo et al. 2015). Por otro lado, en ambas enfermedades se observan fenotipos muy heterogéneos que en algunos casos pueden ser explicados por variaciones genéticas.

En este trabajo nos propusimos identificar variantes genéticas que alterasen el motivo RCGTG de unión de HIF en el genoma de pacientes con la EPOC y el SAOS, con el propósito de estudiar su efecto funcional en la respuesta a hipoxia y analizar su correlación con fenotipos observados en la clínica de ambas enfermedades.

### **6.1 Identificación de variantes genéticas que modifican sitios de unión HIF en el genoma.**

Para poder identificar los polimorfismos que alteran los motivos RCGTG en el genoma en primer lugar definimos las regiones con alta probabilidad de unir HIF. Aunque lo ideal hubiese sido secuenciar por NGS el genoma completo de los pacientes con EPOC y SAOS para identificar la totalidad de las variantes genéticas en motivos RCGTG, los costes de secuenciación de genomas completos aún no son lo suficientemente reducidos como para secuenciar el número de pacientes necesario para los estudios de correlación con fenotipos. Así, con el fin de obtener un listado de regiones que unen HIF hemos utilizado los datos experimentales de ensayos de ChIP-Chip y ChIP-Seq de seis publicaciones en las que se incluyen los sitios de unión de HIF en cinco líneas celulares humanas expuestas a hipoxia. Con esta información podríamos haber elaborado un listado exhaustivo en el que estuvieran representadas todas las regiones identificadas en cada estudio individual. Sin embargo, este listado incluiría un número de regiones demasiado elevado para abordarlo con los medios a nuestro alcance. Por otra parte, una fracción de las regiones incluidas en cada estudio individual pueden ser falsos positivos, de forma que, la suma de todos ellos incluiría un elevado número de regiones espurias. Por estas razones, decidimos centrar el análisis en aquellas regiones que estuvieran representadas en más de un único estudio. Con ésta estrategia aumentamos las posibilidades de secuenciar solo regiones funcionales. Sin embargo, este abordaje elimina los

sitios de unión de HIF específicos de un solo tipo celular. Por lo tanto, todos los resultados de esta tesis, deben entenderse teniendo en cuenta esta limitación.

Mas allá de su importancia para el presente estudio, este listado puede resultar de gran utilidad como conjunto fidedigno de sitios de unión de HIF. En este sentido, es importante destacar que todas las regiones seleccionadas presentan elementos funcionales indicativos de actividad transcripcional de acuerdo a los datos generados por el proyecto ENCODE (Rosenbloom et al. 2013).

El análisis de las regiones seleccionadas para nuestro estudio puso de manifiesto que la mayoría de las regiones que unen HIF corresponden a promotores, en concordancia con lo observado en trabajos previos donde la mayoría de los sitios de unión de HIF se encuentran ubicados a menos 2Kb en 5' del sitio de inicio de la transcripción (TSS) (Mole et al. 2009). En concreto, en dicho trabajo se observa una alta concentración de motivos RCGTG a unas 500pb en 5' del TSS, lo cual se ha visto para otros factores de transcripción, como OCT4, SOX2, NANOG, HNF1A, HNF4A, HNF6, FOXA2, USF1 y CREB1 (Koudritsky & Domany 2008). Por último, se realizó un estudio de ontología de los genes asociados a las regiones reguladoras que unen HIF confirmándose un enriquecimiento en funciones y rutas metabólicas reguladas por hipoxia.

La convergencia de los diferentes tipos de datos incluidos en nuestro análisis nos ha permitido generar un listado robusto de regiones con alta probabilidad de unir HIF, que se localizan en regiones reguladoras de la transcripción (promotores y enhancers) y en proximidad de genes implicados en la respuesta adaptativa a la hipoxia. La secuenciación de estas regiones en 96 pacientes de EPOC y 144 de SAOS, nos permitió identificar un total de 72 variantes localizadas en motivos RCGTG. La presencia de estas variantes es en gran parte producto de sustituciones nucleotídicas en el ADN que pueden ser de dos tipos:

- Transiciones: sustitución de una purina por purina; o de una pirimidina por pirimidina
- Transversiones: sustitución de una purina por pirimidina o viceversa

Si las sustituciones nucleotídicas ocurriesen exclusivamente por azar, existiría el doble de probabilidad de que ocurra una transversión que de que ocurra una transición. Sin embargo, el mecanismo molecular responsable de su generación favorece la formación de transiciones (Collins & Jukes 1994) y de hecho 2/3 de las variantes identificadas son transiciones frente a 1/3 de transversiones (Zhang & Gerstein 2003). En consonancia, el 75% de las variantes identificadas en pacientes de EPOC y SAOS corresponden a transiciones.

Por último, es importante destacar una limitación adicional de nuestro trabajo. Aunque nos hemos centrado en el análisis de variantes que alteran motivos RCGTG, nuestro estudio identificó un número muy elevado de variantes fuera de estos motivos. Estas variantes, podrían afectar la transcripción mediada por HIF sin afectar al HRE, por ejemplo afectando al sitio de unión de factores



de la maquinaria basal o al reclutamiento de otros factores de transcripción que colaboren con HIF. De acuerdo con esta posibilidad, recientemente se describió una variante que afecta a la progresión del carcinoma renal de célula clara al alterar la respuesta transcripcional de HIF sin alterar un motivo RCGTG (Schödel et al. 2012). Por lo tanto, es posible que algunas de las variantes identificadas, pero no estudiadas, puedan alterar la inducción de dianas de HIF y contribuir a la progresión de EPOC y/o SAOS.

## 6.2 Caracterización funcional de SNPs en sitios de unión de HIF.

El análisis funcional de los SNPs que se localizan en el núcleo central del sitio de unión de HIF puso de manifiesto la existencia de variantes genéticas, en el promotor de *PGK1* y en un intrón de *EGFR*, que se encuentran en regiones no codificantes del genoma y disminuyen la inducción transcripcional de genes en hipoxia, posiblemente debido a una pérdida de unión del factor a su secuencia HRE.

Este resultado corrobora la importancia de la integridad del motivo RCGTG en la modulación transcripcional mediada por HIF, previamente observado en un trabajo de nuestro laboratorio (Ortiz-Barahona et al. 2010). Además, puesto que estos SNPs han sido identificados en pacientes con EPOC y SAOS, se pone de manifiesto la existencia de variantes genéticas que podrían tener efectos funcionales en individuos con enfermedades que cursan con hipoxia.

En un estudio en el que se realizó ChIP-Seq para el factor nuclear  $\kappa$ B (p65) en muestras humanas, se ha observado que las diferencias en la unión de este factor se asocian a SNPs que provocan diferencias en la expresión génica (Kasowski et al. 2010). Este resultado sugiere que los efectos diferenciales observados en la transcripción son consecuencia de la variación genética en el sitio de unión del factor de transcripción. En otro estudio en el que se analiza como los polimorfismos afectan los niveles de expresión génica, se evalúan SNPs en regiones no codificantes cercanas a *CHI3L2* y *CRYZ* y se analizan sus transcriptos en experimentos de RNA-Seq observando que los niveles de expresión génica en células B varían en función del genotipo de los individuos (Cheung et al. 2010). Por lo que la variación genética en regiones no codificantes modula la expresión génica de manera diferencial.

Uno de los SNPs que hemos caracterizado destruye un motivo RCGTG en el promotor del gen de *PGK1*, rs150921338. *PGK1* es una enzima glucolítica esencial para la síntesis de ATP, cuya regulación transcripcional en hipoxia ha sido descrita previamente (Semenza et al. 1994; Semenza et al. 1996). Nuestros resultados indican que el SNP rs150921338 en *PGK1* disminuye la inducción génica en hipoxia. El efecto parcial de este SNP en la inducción de *PGK1* puede explicarse por la existencia de otro HRE en su promotor (Semenza et al. 1994; Semenza et al. 1996). Dada esta redundancia en la regulación de *PGK1*, es difícil predecir la implicación del polimorfismo en los pacientes puesto que sus efectos podrían estar compensados. También cabe la posibilidad de que cada uno de los HREs

tenga un papel predominante en diferentes tejidos, de forma que el control de la expresión de PGK1 este mediado por uno u otro en diferentes tipos celulares. En este caso, la presencia del SNP rs150921338 en homocigosis, podría contribuir a la progresión de las enfermedades obstructivas pulmonares al alterar la respuesta glicolítica de algunos tejidos en hipoxia.

Por otro lado, el SNP rs6593210 identificado en el locus de *EGFR*, suprime por completo la inducción de la transcripción de este gen en hipoxia. Este efecto sumado a que la activación de *EGFR* implica la regulación de múltiples respuestas celulares como proliferación, diferenciación, motilidad y supervivencia (Jorissen 2003) plantea la posibilidad de que éste SNP tenga consecuencias severas en la respuesta a hipoxia, sobre todo en aquellos pacientes que lo presenten en homocigosis.

La regulación de la expresión de *EGFR* en hipoxia ha sido descrita previamente por Franovic *et al.* en un trabajo en el que determinan que su regulación es exclusivamente a nivel traduccional, pues observan el aumento de EGFR a nivel proteico en células MDA-MB-231, PC3 y U87MG, pero al estudiar los niveles de ARNm no observaron cambios por PCR-RT (Franovic et al. 2007). Sin embargo, nuestros resultados demuestran que existe un componente de regulación transcripcional ya que al analizar por qPCR-RT los niveles de mensajero de EGFR en los experimentos de expresión genica de alta resolución hemos observado un aumento de la síntesis del ARNm de *EGFR* en hipoxia en células HeLa. La diferencia entre nuestros resultados y los de Franovic *et al.* pueden deberse por un lado, a que EGFR no este regulado a nivel transcripcional en todas las líneas celulares, además de que en nuestro trabajo hemos usado líneas celulares diferentes a las usadas en el trabajo de Franovic. Por otro lado, en dicho trabajo miden los niveles de mensajero por PCR-RT a tiempo final que es una técnica que no es cuantitativa y ello se suma a que la regulación por hipoxia del mensajero de EGFR es apenas de dos veces lo cual dificulta la determinación de esta inducción por PCR-RT.

La identificación del SNP rs6593210 que afecta la expresión de *EGFR* en hipoxia resulta de especial interés en los pacientes con EPOC, pues se ha descrito que la activación de este receptor es responsable de un aumento de la expresión de MUC5AC, la mucina predominante en el epitelio de las vías respiratorias, (Takeyama et al. 1999). Por tanto, la presencia de éste SNP podría modificar el pronóstico de los pacientes que presentan bronquitis crónica que se caracteriza por la hipersecreción de mucus. De hecho, se describió que la expresión de *EGFR* en el epitelio de las vías aéreas se encuentra aumentado en pacientes con EPOC respecto de sujetos sanos (O'Donnell et al. 2004; Boer et al. 2006).

Por otra parte, el aumento de la secreción de mucus en pacientes con EPOC se ha descrito en otro trabajo donde se ha observado el aumento de las células caliciformes productoras de mucina en las vías aéreas pequeñas en análisis histopatológicos de estos individuos (Burgel & Nadel 2008). Además, también se ha descrito que el humo del tabaco, que es el principal factor de riesgo en la EPOC,

provoca hipersecreción de la mucina MUC5AC en las vías aéreas mediante el aumento del ARNm de *EGFR* (Takeyama et al. 2001).

Dada la importancia de EGFR, se ha evaluado como herramienta terapéutica para reducir el mucus en algunos modelos preclínicos de EPOC, mediante la inhibición de su actividad tirosina quinasa. Se ha descrito que el inhibidor de EGFR AG1478 en células epiteliales bronquiales humanas atenúa la expresión de MUC5AC y la activación de ERK (Li et al. 2012). En otro trabajo en el que también usan AG1478 y otro inhibidor llamado BIBX1522 se bloquea por completo la activación de EGFR, así como la inducción del ARNm de MUC5AC y su proteína en células de carcinoma mucoepidermoide del pulmón (Kohri et al. 2002). Estos trabajos demuestran que EGFR es una diana terapéutica relevante para reducir la producción de mucus y sugieren que el eje EGFR-MUC5AC puede contribuir de forma significativa a la progresión de pacientes de EPOC y, por tanto, es posible que la presencia del SNP rs6593210 en homocigosis pueda alterar este aspecto de la enfermedad.

En los pacientes con EPOC que presentan este SNP rs6593210 no se observaron diferencias en la aparición de la bronquitis crónica ni en la gravedad de la enfermedad, comparado con los individuos que no presentan el SNP. Sin embargo, el conjunto de todas estas evidencias sugieren que la presencia del SNP rs6593210 podría suponer un mejor pronóstico en aquellos pacientes con EPOC que presenten bronquitis crónica, pues una menor expresión de *EGFR* provocaría una menor secreción de mucus en las vías aéreas pequeñas facilitando el intercambio gaseoso. En este trabajo no hemos estudiado la producción de mucus lo que resulta de especial interés en vista de estos resultados.

Recientemente, se ha identificado otro SNP en regiones no codificantes del gen de *EGFR* en pacientes con EPOC (Zhou et al. 2015). Este polimorfismo 774 T>C se encuentra en la región 3' no traducida (del inglés *untranslated region*, UTR) de EGFR modificando el sitio de unión del microARN miR-214, de manera que los pacientes que presentan el genotipo CC tienen aumentados los niveles de EGFR en células de músculo liso arteriales (PASMC). Además este SNP se asocia a un riesgo mayor de padecer hipertensión pulmonar en pacientes con EPOC (OR: 1,965; p= 0.0095). Este trabajo muestra otra vía de regulación de EGFR donde la variación genética en regiones no codificantes juega un papel central.

### **6.3 Edición del genoma como herramienta para el análisis funcional de las variantes genéticas en regiones reguladoras.**

El sistema CRISPR/Cas 9 nos ha sido de gran utilidad para validar el efecto del SNP rs6593210 en *EGFR* en su regulación transcripcional por hipoxia; permitiéndonos demostrar que la presencia del SNP disminuye la inducción transcripcional de *EGFR* en células HEK293T. En paralelo, como control positivo se mutó el HRE funcional del enhancer de *EGLN3* observando una disminución de la transcripción en hipoxia en la misma línea celular. Sin embargo, el efecto del SNP de *EGFR* en

homocigosis fue de menor magnitud que el observado para *EGLN3* en heterocigosis, lo cual se debe a que la inducción de *EGFR* en hipoxia es de sólo dos veces mientras que la de *EGLN3* es mayor. Además, existe una cierta variabilidad en la respuesta a hipoxia incluso en clones del mismo genotipo; por lo que sería importante reforzar los resultados obtenidos utilizando un número mayor de clones editados.

Estos resultados demuestran el papel de las variantes genéticas que afectan a motivos de unión de HIF. Recientemente, el grupo de Wenger consiguió un resultado similar mediante edición génica de un HRE utilizando TALENs. En este trabajo se modificó el sitio de unión de HIF localizado a 82 KB en 5' del TSS de *PAG1*, observando la pérdida de la inducción transcripcional de *PAG1* en hipoxia (Schorg et al. 2015). Además, se demostró la pérdida de la interacción entre el HRE editado vía TALENs y el promotor de *PAG1* mediante Captura de Conformación del Cromosoma (3C) en normoxia e hipoxia en células en las que se alteró el sitio de unión de HIF.

Inicialmente cuando nos propusimos utilizar el sistema CRISPR/Cas9 para editar el genoma pensamos en analizar un mayor número de SNPs. Sin embargo, el número de clones editados por HDR resultó ser muy pequeño en los tres locus en los que hemos trabajado. Por ello, es imprescindible mejorar la eficiencia de la edición génica mediante HDR, para seguir adelante con la caracterización funcional de otros SNPs.

La utilización del sistema CRISPR/Cas9 para editar el genoma vía HDR ha sido menos frecuentemente utilizada que la reparación vía NHEJ. En los trabajos publicados hasta la fecha se pone en evidencia la baja eficiencia de la HDR en múltiples organismos modificados como *Drosophila* (Zhang et al. 2014), peces cebra (Hruscha et al. 2013) y ratones (Maruyama et al. 2015). También se observan pequeños porcentajes de HDR en células humanas que oscilan entre un 1-20% en HEK293T y HUES9 (Lin et al. 2014; Ran et al. 2013).

Esta baja en eficiencia en la HDR en células de mamíferos es en parte debida a que los daños de doble cadena del ADN son mayormente reparados mediante la NHEJ en detrimento de la HDR (Mao et al. 2008; Lieber 2010; Li et al. 2014). Por ello, actualmente se están desarrollando estrategias que aumentan el porcentaje de HDR, como la búsqueda sistemática de compuestos de bajo peso molecular que activen esta vía de reparación del ADN. De esta forma, se ha identificado que el L755507 y la Brefeldina A aumentan la eficiencia de la HDR 3 veces para las inserciones de fragmentos largos y en 9 veces para el caso de mutaciones puntuales en células madre embrionarias de ratones (Yu et al. 2015). Por otro lado, Chu *et al.* ha logrado mejorar la eficiencia de HDR en 4- 5 veces silenciando moléculas claves en el proceso de NHEJ como KU70 o ADN ligasa IV (Chu et al. 2015). Otra opción es el uso de inhibidores de la NHEJ, como Scr7 que aumenta la eficiencia de HDR hasta 19 veces en células de mamíferos y en ratones (Maruyama et al. 2015). Por último, otra

estrategia mediante la que se ha logrado potenciar la HDR frente a NHEJ, es modificando el ADN donador; observándose que el donador de cadena sencilla, frente a ADN de doble cadena, así como el aumento de su longitud, mejoran la eficiencia de la HDR (Lin et al. 2014). Adicionalmente, en el mismo trabajo, se ha demostrado que la fase del ciclo celular afecta a la eficiencia de la HDR.

En un futuro próximo estamos interesados en explorar la utilidad de estas estrategias para aumentar la eficiencia de la edición de SNPs en regiones de unión de HIF.

#### **6.4 Correlación entre las variantes genéticas en sitios de unión de HIF y los fenotipos de la EPOC y el SAOS.**

Dada la gran heterogeneidad fenotípica que se observa en pacientes con EPOC y SAOS (Berndt et al. 2012; Casale et al. 2009) y puesto que en ambas enfermedades la activación de HIF es central en la regulación de la expresión génica, las variantes genéticas que afectan HREs podrían contribuir a explicar los fenotipos o comorbilidades de EPOC y SAOS.

En los pacientes con SAOS, se identificaron dos SNPs (rs1730190 y rs1657290) para los que observamos una tendencia a la correlación con una menor probabilidad de padecer HTA. Por ello, sería necesario realizar este análisis en una cohorte independiente de pacientes para confirmar esta asociación. Los genes más próximos a estos SNPs son *HTR5A* e *INSIG1*, que se encuentran a 79 Kb en su extremo 5' y a 150 KB hacia el extremo 3', respectivamente. *INSIG1* es un gen del que tenemos evidencia de su inducción en hipoxia en el meta-análisis de perfil de expresión génica en hipoxia realizado previamente en el laboratorio (Ortiz-Barahona et al. 2010). Sin embargo, para *HTR5A* no se observa inducción en dicho meta-análisis. *HTR5A* es un receptor de serotonina acoplado a proteínas G que se asocia a la regulación del sueño para el que ya existe una asociación con el riesgo aumentado de SAOS, comentado previamente en la introducción (Zhao et al. 2013), pero esta asociación se corresponde con un SNP más cercano al gen *HTR5A* (-1438) diferente a los que hemos identificado. Previamente, se ha descrito que los receptores *HTR5A* se encuentran en zonas quimiorreceptoras de la aorta de cerdos pudiendo mediar la respuesta ventilatoria (Serra et al. 2002). Por tanto, confirmar la asociación de los SNPs rs1730190 y rs1657290 con HTA en pacientes de SAOS resulta interesante.

El encontrar una variante genética asociada a una menor susceptibilidad de padecer HTA puede ser de utilidad puesto que la hipertensión es un síntoma estrechamente asociado a SAOS. Los mecanismos involucrados en el desarrollo de la HTA son conocidos e implican un aumento de la actividad quimiorreceptora del cuerpo carotideo por la hipoxia intermitente, que conlleva un aumento en la actividad simpática, provocando el aumento en la secreción de catecolaminas que aumentan el tono arterial (Prabhakar et al. 2007). Por otro lado, los radicales libres generados por la

isquemia y reperfusión asociados a las apneas, provocan estrés oxidativo que conduce a la lesión del endotelio con su consecuente disfunción (Cortada et al. 2009).

La asociación entre SAOS y HTA ha sido previamente estudiada, en varios trabajos en los que se determina que existe una relación dosis-respuesta entre la gravedad de SAOS y la HTA que presentan los pacientes (Hla et al. 1994; Peppard et al. 2000; García-Río 2000). Sin embargo, nuestro trabajo identifica por primera vez polimorfismos que afectan el motivo de unión a HIF y su posible asociación con HTA en pacientes con SAOS.

El análisis de la correlación genotipo-fenotipo en los pacientes con EPOC nos ha permitido poner de manifiesto la asociación de la variante chr20:50.366.859:A, SNP rs118151281 con una mayor probabilidad de padecer HTA y RCV. La HTA es uno de los principales factores que influyen en el riesgo cardiovascular en pacientes con la EPOC (de Lucas-Ramos et al. 2008), por ello es posible que la asociación del SNP rs118151281 con ambas comorbilidades estén relacionadas entre sí.

El SNP rs118151281 para el que observamos asociación en los pacientes con EPOC se encuentra en el primer intrón de *ATP9A*, aunque no hemos logrado asociar a qué gen estaría regulando el motivo RCGTG en que se encuentra ni caracterizar éste SNP funcionalmente, hemos encontrado que este SNP se encuentra en una región de 28Mb asociada a hipertensión (OR=4,24; p = 0,002)(Wallace et al. 2006), lo que refuerza la asociación encontrada con HTA. La alternativa más conveniente para determinar la identidad del gen regulado por el motivo RCGTG sería realizar RNA-Seq de una línea celular en la que se ha introducido el SNP, de manera que se espera encontrar un grupo de genes regulados diferencialmente en hipoxia en la línea editada respecto de la línea celular de referencia y así poder identificar un conjunto acotado de genes donde centrar el estudio funcional. Por otro lado, para estudiar los efectos funcionales del SNP se podrían realizar ensayos de luciferasa en otro vector reportero donde no se observe inducción basal en hipoxia. Otra opción, es realizar experimentos de ChIP para HIF en una línea celular en la que se ha introducido el SNP en el genoma. De modo que, en caso de ser un motivo funcional, esperaríamos que se pierda la unión del factor de transcripción en dicha región respecto de la línea celular de referencia.

Dada la multiplicidad de factores genéticos implicados en la EPOC, ha surgido COPDGene que es un estudio observacional multicéntrico diseñado para identificar los factores genéticos asociados a la EPOC y que tiene como último fin el poder interpretar la heterogeneidad individual en la susceptibilidad y la respuesta a tratamientos de los pacientes con esta enfermedad (Regan et al. 2011). A partir de datos de éste proyecto se han identificado tres locus (CHRNA3, FAM13A, HHIP), asociados a estadios de mayor gravedad de la enfermedad en un estudio de asociación de todo el genoma en miles de pacientes (Cho et al. 2012). Recientemente, en otro estudio se han identificado

dos polimorfismos en regiones intrónicas de FAM13A, las variantes rs1903003 y rs7671167 que confieren un efecto protector respecto a la EPOC (ambos  $OR < 0,405$ ;  $P < 0,002$ ) (Ziółkowska-Suchanek et al. 2015).

Hasta el momento, la asociación entre el SNP rs118151281 con HTA y RCV constituye la primera evidencia en la que se pone de manifiesto los efectos de la variación genética en el sitio de unión de HIF y sus consecuencias en el fenotipo en EPOC.

Las asociaciones que hemos encontrado son con SNPs de variantes comunes (MAF mayores al 5%), y en general se ha visto que este tipo de variantes suelen presentar efectos fenotípicos modestos mientras que variantes raras se asocian a efectos fenotípicos más profundos (Gorlov et al. 2008). Esta podría ser una razón por la cual no hemos encontrado asociaciones con los SNPs que localizan en *PGK1* y *EGFR* que hemos caracterizado en ensayos funcionales.

Actualmente, se están realizando estudios a gran escala en los que se analiza el efecto de diversas variantes genéticas y el fenotipo, considerando la contribución de factores ambientales; ya que existen evidencias de que la combinación de alteraciones en varios genes con efectos débiles, sumado a la contribución de factores ambientales es lo que puede tener un impacto significativo en el fenotipo pues aporta información que se pierde cuando se analizan solo causas genéticas aisladas (Chen et al. 2013). Un estudio de este tipo ha analizado diversas comorbilidades que sufren los pacientes de EPOC identificando que muchas de ellas se relacionan a nivel biológico y molecular a través de genes, proteínas y rutas biológicas que son compartidas (Grosdidier et al. 2014), manifestando el carácter génico multifactorial de la EPOC así como la importancia de abordar el estudio de esta enfermedad a nivel global.

En conjunto, en este trabajo hemos puesto de manifiesto por primera vez la importancia de la variación genética en sitios de unión de HIF en pacientes con EPOC y SAOS. Hemos identificado SNPs que afectan la regulación de la expresión génica en condiciones de hipoxia. Además, hemos encontrado que existe una asociación de algunos de los SNPs en motivos RCGTG identificados y algunas comorbilidades relevantes de ambas enfermedades. Estos resultados, sugieren que las variaciones genéticas en sitios de unión de HIF podrían contribuir a explicar parte de la heterogeneidad observada en los pacientes que padecen estas enfermedades.





## **7 CONCLUSIONES**



1. Mediante la integración de datos experimentales sobre los sitios de unión de HIF a nivel genómico, perfiles de expresión en hipoxia e información sobre elementos funcionales generada en el proyecto ENCODE, generamos un listado fidedigno de sitios de unión de HIF a nivel genómico.
2. Mediante secuenciación masiva hemos identificado 72 variantes de un único nucleótido que alteran motivos RCGTG, en regiones que unen HIF en el genoma de pacientes con Enfermedad Pulmonar Obstructiva Crónica (EPOC) y Síndrome de Apnea Obstructiva del Sueño (SAOS).
3. El análisis de polimorfismos de un único nucleótido que destruyen motivos RCGTG, mediante ensayos reporteros y la edición de genoma, nos ha permitido validar el efecto funcional de varios de los SNPs identificados. En concreto, el SNP, rs150921338, localizado en el promotor de *PGK1*, y el SNP, rs6593210, en un intrón de *EGFR*, alteran la inducción transcripcional de estos genes en hipoxia.
4. El análisis preliminar de la asociación entre las variantes genéticas encontradas en motivos RCGTG y los fenotipos o comorbilidades de EPOC y SAOS sugiere una asociación entre el SNP rs118151281 y una mayor probabilidad de padecer hipertensión arterial y/o riesgo cardiovascular en pacientes con EPOC. Por otro lado, en los pacientes con SAOS se evidencia una tendencia de asociación entre el SNP rs1730190 y rs1657290 con una menor probabilidad de padecer hipertensión arterial.



## **8 BIBLIOGRAFÍA**



- American Diabetes Association., 2013. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, 36, pp.581–90.
- Aragones, J. et al., 2001. Evidence for the Involvement of Diacylglycerol Kinase in the Activation of Hypoxia-inducible Transcription Factor 1 by Low Oxygen Tension. *Journal of Biological Chemistry*, 276(13), pp.10548–10555.
- Aragonés, J. et al., 2008. Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism. *Nature Genetics*, 40(2), pp.170–180.
- Artigas, M.S. et al., 2011. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nature Genetics*, 43(11), pp.1082–1090.
- Bartels, N.K. et al., 2007. Risk factors and myocardial infarction in patients with obstructive sleep apnea: impact of beta2-adrenergic receptor polymorphisms. *BMC medicine*, 5(1).
- Beck, A.T. et al., 1961. An inventory for measuring depression. *Archives of General Psychiatry*, 4, pp.561–571.
- Berndt, A., Leme, A.S. & Shapiro, S.D., 2012. Emerging genetics of COPD. *EMBO Molecular Medicine*, 4(11), pp.1144–1155.
- Boer, W.I. De et al., 2006. Expression of Epidermal Growth Factors and Their Receptors in the Bronchial Epithelium of Subjects With Chronic Obstructive Pulmonary Disease. *American Journal of Clinical Pathology*, 125(2), pp.184–192.
- Brusselmans, K. et al., 2003. Heterozygous deficiency of hypoxia-inducible factor-2 $\alpha$  protects mice against pulmonary hypertension and right ventricular dysfunction during prolonged hypoxia. *Journal of Clinical Investigation*, 111(10), pp.1519–1527.
- Burgel, P.-R. & Nadel, J. a, 2008. Epidermal growth factor receptor-mediated innate immune responses and their roles in airway diseases. *European Respiratory Journal*, 32(4), pp.1068–1081.
- Cai, Z. et al., 2007. Complete loss of ischaemic preconditioning-induced cardioprotection in mice with partial deficiency of HIF-1. *Cardiovascular Research*, 77(3), pp.463–470.
- Calle, M. et al., 2012. Guía Española de la EPOC ( GesEPOC ). Tratamiento farmacológico de la EPOC. *Archivos de bronconeumología*, 48(7), pp.247–257.
- Cao, C. et al., 2014. Vascular Endothelial Growth Factor Genotypes and Haplotypes Contribute to the Susceptibility of Obstructive Sleep Apnea Syndrome. *PLoS ONE*, 9(12), p.e114582.
- Carroll, D., 2014. Genome Engineering with Targetable Nucleases. *Annual Review of Biochemistry*, 83(1), pp.409–439.
- Casale, M. et al., 2009. Obstructive sleep apnea syndrome: from phenotype to genetic basis. *Current Genomics*, 10(2), pp.119–126.
- Celli, B.R. et al., 2004. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *European Respiratory Journal*, 23(6), pp.932–946.

- Celli, B.R. et al., 2004. The Body-Mass Index, Airflow Obstruction, Dyspnea, and Exercise Capacity Index in Chronic Obstructive Pulmonary Disease. *The New England Journal of Medicine*, 350(10), pp.1005–1012.
- Cerón, E.M., Mateos, C. & García-río, F., 2015. Sleep Apnea – Hypopnea Syndrome and Type 2 Diabetes . A Reciprocal Relationship ? *Archivos de Bronconeumología*, 51(3), pp.128–139.
- Chen, X., Xu, X. & Xiao, F., 2013. Heterogeneity of chronic obstructive pulmonary disease: from phenotype to genotype. *Frontiers of Medicine*, 7(4), pp.425–432.
- Cheung, V.G. et al., 2010. Polymorphic cis- and Trans-Regulation of human gene expression. *PLoS Biology*, 8(9).
- Cho, M.H., McDonald, M.-L.N. & Zhou, X., 2012. Risk loci for chronic obstructive pulmonary disease: a genomewide association study and meta-analysis. *Lancet Respir Med*, 127(3), pp.214–225.
- Chu, V.T. et al., 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, 33(5), pp.543–548.
- Cipolleschi, M.G., Dello Sbarba, P. & Olivotto, M., 1993. The role of hypoxia in the maintenance of hematopoietic stem cells. *Blood*, 82(7), pp.2031–2037.
- Collins, D.W. & Jukes, T.H., 1994. Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics*, 20(3), pp.386–396.
- Cortada, G.T. et al., 2009. Síndrome de apneas-hipopneas del sueño , hipertensión y riesgo cardiovascular. *Hipertensión y Riesgo Vascular*, 26(2), pp.72–79.
- Crapo, R.O. et al., 2002. ATS statement: Guidelines for the six-minute walk test. *American Journal of Respiratory and Critical Care Medicine*, 166(1), pp.111–117.
- Divo, M. et al., 2012. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.*, 186(2), pp.155–61.
- Dölken, L. et al., 2008. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA Journal*, 14, pp.1959–1972.
- Ebert, B.L. & Bunn, H.F., 1998. Regulation of transcription by hypoxia requires a multiprotein complex that includes hypoxia-inducible factor 1, an adjacent transcription factor, and p300/CREB binding protein. *Molecular and Cellular Biology*, 18(7), pp.4089–4096.
- Ehrismann, D. et al., 2007. Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *The Biochemical Journal*, 401(1), pp.227–234.
- Enattah, N.S. et al., 2002. Identification of a variant associated with adult-type hypolactasia. *Nature genetics*, 30(2), pp.233–237.
- Ernst, J. et al., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345), pp.43–49.



- Ferrer, M. et al., 1996. Validity and reliability of the St George's respiratory questionnaire after adaptation to a different language and culture: The Spanish example. *European Respiratory Journal*, 9, pp.1160–1166.
- Findley, L. et al., 1986. Cognitive impairment in patients with obstructive sleep apnea and associated hypoxemia. *Chest*, 90(5), pp.686–690.
- Firth, J.D., Ebert, B.L. & Ratcliffe, P.J., 1995. Hypoxic Regulation of Lactate Dehydrogenase A. *Journal of Biological Chemistry*, 270(36), pp.21021–21027.
- Franovic, A. et al., 2007. Translational up-regulation of the EGFR by tumor hypoxia provides a nonmutational explanation for its overexpression in human cancer. *PNAS*, 104(32), pp.13092–13097.
- Freedman, S.J. et al., 2003. Structural basis for negative regulation of hypoxia-inducible factor-1 $\alpha$  by CITED2. *Nature Structural Biology*, 10(7), pp.504–512.
- Galbraith, M.D. et al., 2013. XHIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell*, 153(6), pp.1327–1339.
- Ganrot, P., Laurell, C. & Eriksson, S., 1967. Obstructive lung disease and trypsin inhibitors in alpha-1-antitrypsin deficiency. *Scand J Clin Lab Invest.*, 19(3), pp.205–208.
- Gao, L. et al., 2014. Glucose sensing by carotid body glomus cells: potential implications in disease. *Frontiers in Physiology*, 5, pp.1–9.
- García-Río, F., 2000. Sleep Apnea and Hypertension. The Role of Peripheral Chemoreceptors and the Sympathetic System. *CHEST Journal*, 117(5), p.1417.
- Global Strategy for the Diagnosis, Management and Prevention of COPD, G.I. for C.O.L.D. (GOLD) 2015., Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015. <http://www.goldcopd.org/>.
- Gorlov, I.P. et al., 2008. Shifting Paradigm of Association Studies: Value of Rare Single-Nucleotide Polymorphisms. *American Journal of Human Genetics*, 82(1), pp.100–112.
- Gozal, D. & Kheirandish-Gozal, L., 2008. Cardiovascular morbidity in obstructive sleep apnea: Oxidative stress, inflammation, and much more. *American Journal of Respiratory and Critical Care Medicine*, 177(4), pp.369–375.
- Grosdidier, S. et al., 2014. Network medicine analysis of COPD multimorbidities. *Respiratory Research*, 15(1), p.111.
- Grungy, S. et al., 2004. Definition of metabolic syndrome : report of the National Heart , Lung , and Blood Institute / American Heart Association Conference on scientific issues related to definition. *Circulation*, 109, pp.433–8.
- Halbert, R.J. et al., 2006. Global burden of COPD: Systematic review and meta-analysis. *European Respiratory Journal*, 28(3), pp.523–532.

- Hale, L.P. et al., 2002. Hypoxia in the thymus: role of oxygen tension in thymocyte survival. *American Journal of Physiology - Heart and Circulatory Physiology*, 282(4), pp.H1467–H1477.
- Hirsilä, M. et al., 2003. Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *Journal of Biological Chemistry*, 278(33), pp.30772–30780.
- Hla, K. et al., 1994. Sleep Apnea and Hypertension: A Population-based Study. *Ann Intern Med*, 120(5), pp.382–8.
- Hoffman, E.C. et al., 1991. Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science*, 252(5008), pp.954–958.
- Hruscha, A. et al., 2013. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development*, 140(24), pp.4982–7.
- Huang, D.W., Sherman, B.T. & Lempicki, R. a, 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), pp.44–57.
- Hunninghake, G.M., Cho, M.H. & Tesfaigzi, Y., 2009. MMP12, Lung Function, and COPD in High-Risk Populations. *New England Journal of Medicine*, 361(27), pp.2599–2608.
- Hunt, K. a et al., 2008. Newly identified genetic risk variants for celiac disease related to the immune response. *Nature Genetics*, 40(4), pp.395–402.
- Iyer, N. V, Kotch, L.E., et al., 1998. Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1alpha. *Genes & Development*, 12(2), pp.149–162.
- Iyer, N. V, Leung, S.W. & Semenza, G.L., 1998. The human hypoxia-inducible factor 1alpha gene: HIF1A structure and evolutionary conservation. *Genomics*, 52(2), pp.159–165.
- Jiang, B.-H. et al., 1997. Transactivation and Inhibitory Domains of Hypoxia-inducible Factor 1 : MODULATION OF TRANSCRIPTIONAL ACTIVITY BY OXYGEN TENSION. *Journal of Biological Chemistry*, 272(31), pp.19253–19260.
- Jiang, F. et al., 2015. A Cas9 – guide RNA complex preorganized for target DNA recognition. *Science*, 348(6242), pp.1477–1481.
- Jorissen, R., 2003. Epidermal growth factor receptor: mechanisms of activation and signalling. *Experimental Cell Research*, 284(1), pp.31–53.
- Jormsjö, S. et al., 2001. Allele-specific regulation of matrix metalloproteinase-7 promoter activity is associated with coronary artery luminal dimensions among hypercholesterolemic patients. *Arteriosclerosis, thrombosis, and vascular biology*, 21(11), pp.1834–1839.
- Kasowski, M. et al., 2010. Variation in transcription factor binding among humans. *Science*, 328(5975), pp.232–235.
- Kent, B.D., Mitchell, P.D. & McNicholas, W.T., 2011. Hypoxemia in patients with COPD: Cause, effects, and disease progression. *International Journal of COPD*, 6(1), pp.199–208.

- Kim, H. & Kim, J., 2014. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics*, 15(5), pp.321–334.
- Kohri, K. et al., 2002. Pseudomonas aeruginosa induces MUC5AC production via epidermal growth factor receptor. *European Respiratory Journal*, 20(5), pp.1263–1270.
- Koudritsky, M. & Domany, E., 2008. Positional distribution of human transcription factor binding sites. *Nucleic Acids Research*, 36(21), pp.6795–6805.
- Lakhdar, R. et al., 2011. Update in chronic obstructive pulmonary disease: role of antioxidant and metabolizing gene polymorphisms. *Exp Lung Res*, 37(6), pp.364–75.
- Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), p.R25.
- Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), pp.2987–2993.
- Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.
- Li, K. et al., 2014. Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System K. Yu, ed. *PLoS ONE*, 9(8), pp.1–10.
- Li, N. et al., 2012. Chronic mechanical stress induces mucin5AC expression in human bronchial epithelial cells through ERK dependent pathways. *Molecular Biology Reports*, 39(2), pp.1019–1028.
- Lieber, M.R., 2010. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annu Rev Biochem.*, 79, pp.181–211.
- Lin, S. et al., 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*, 3, pp.1–13.
- Lloberes, P. et al., 2011. Diagnóstico y tratamiento del síndrome de apneas-hipopneas del sueño. *Archivos de Bronconeumología*, 47(3), pp.143–156.
- Loenarz, C. et al., 2011. The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, *Trichoplax adhaerens*. *EMBO reports*, 12(1), pp.63–70.
- De Lucas-Ramos, P. et al., 2008. Cardiovascular Risk Factors in Chronic Obstructive Pulmonary Disease: Results of the ARCE Study. *Archivos de Bronconeumología*, 44(5), pp.233–238.
- MacIntyre, N. et al., 2005. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *European Respiratory Journal*, 26(4), pp.720–735.
- Maes, C., Carmeliet, G. & Schipani, E., 2012. Hypoxia-driven pathways in bone development, regeneration and disease. *Nature Reviews Rheumatology*, 8, pp.358–366.

- Mahler, D.A., 1987. Dyspnea: diagnosis and management. *Clin Chest Med.* 1987 Jun;8(2):215-30., 8(2), pp.215–30.
- Makino, Y. et al., 2001. Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression. *Nature*, 414(6863), pp.550–4.
- Mali, P. et al., 2013. RNA-Guided Human Genome Engineering via Cas9. *Science*, 339(6121), pp.823–826.
- Mancia, G. et al., 2007. 2007 ESH-ESC Practice Guidelines for the Management of Arterial Hypertension: ESH-ESC Task Force on the Management of Arterial Hypertension. *J Hypertens.* 2007 Sep;25(9):1751-62., 25(9), pp.1751–62.
- Mannino, D. & Buist, A., 2007. Global burden of COPD: risk factors, prevalence, and future trends. *Lancet*, 370(9589), pp.765–73.
- Mannino, D.M. et al., 2006. The natural history of chronic obstructive pulmonary disease. *European Respiratory Journal*, 27(3), pp.627–643.
- Mao, Z. et al., 2008. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair*, 7(10), pp.1765–1771.
- Marin, J. et al., 2005. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *Lancet*, 365(9464), pp.1046–53.
- Marin, J. et al., 1997. Prevalence of sleep apnoea syndrome in the Spanish adult population. *International Journal of Epidemiology*, 26(2), pp.381–386.
- Maruyama, T. et al., 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology*, 33(5), pp.1–9.
- Mathelier, A., Shi, W. & Wasserman, W.W., 2015. Identification of altered cis -regulatory elements in human disease. *Trends in Genetics*, 31(2), pp.67–76.
- Maxwell, P.H. et al., 1999. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature*, 399(6733), pp.271–275.
- Mcgrady, T. et al., 2015. Journal of the COPD Foundation Chronic Obstructive Pulmonary Diseases : Characteristics of Chronic Obstructive Pulmonary Disease ( COPD ) Patients Reporting Alpha-1 Antitrypsin Deficiency in the WebMD Lung Health Check Database. , 2(2), pp.141–151.
- McNicholas, W., 2008. Diagnosis of obstructive sleep apnea in adults. *Proc Am Thorac Soc.*, 5(2), pp.154–60.
- McVicker, G. et al., 2013. Identification of Genetic Variants That Affect Histone Modifications in Human Cells. *Science*, 342(6159), pp.747–749.
- De Miguel Díez, J., García, T.G. & Maestu, L.P., 2010. Comorbilidades de la EPOC. *Archivos de Bronconeumología*, 46(11), pp.20–25.

- Miller, M.R. et al., 2005. Standardisation of spirometry. *European Respiratory Journal*, 26(2), pp.319–338.
- Miravittles, M. et al., 2013. Characterisation of the overlap COPD-asthma phenotype. Focus on physical activity and health status. *Respiratory Medicine*, 107(7), pp.1053–1060.
- Miravittles, M. et al., 2009. Prevalence of COPD in Spain: impact of undiagnosed COPD on quality of life and daily life activities. *Thorax*, 64(10), pp.863–868.
- Miravittles, M. et al., 2014. Spanish Guideline for COPD (GesEPOC). Update 2014. *Archivos de Bronconeumología*, 50(S1), pp.1–16.
- Moffatt, M.F. et al., 2007. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*, 448(7152), pp.470–473.
- Mole, D.R. et al., 2009. Genome-wide Association of Hypoxia-inducible Factor (HIF)-1 and HIF-2 DNA Binding with Expression Profiling of Hypoxia-inducible Transcripts. *Journal of Biological Chemistry*, 284(25), pp.16767–16775.
- Morley, M. et al., 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature*, 430(7001), pp.743–747.
- Myers, R.M. et al., 2011. A User's Guide to the Encyclopedia of DNA Elements (ENCODE) P. B. Becker, ed. *PLoS Biology*, 9(4), p.e1001046.
- Nanduri, J. et al., 2015. Hypoxia-inducible factors and hypertension: lessons from sleep apnea syndrome. *Journal of Molecular Medicine*, 93(5), pp.473–480.
- O'Donnell, R. a et al., 2004. Expression of ErbB receptors and mucins in the airways of long term current smokers. *Thorax*, 59(12), pp.1032–1040.
- Ortiz-Barahona, A. et al., 2010. Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and in silico binding site prediction. *Nucleic Acids Research*, 38(7), pp.2332–2345.
- Özdaş, S. et al., 2015. Association of Interleukin-10 gene promoter polymorphisms with obstructive sleep apnea. *Sleep and Breathing*, in press.
- Peng, Y.-J. et al., 2006. Heterozygous HIF-1 $\alpha$  deficiency impairs carotid body-mediated systemic responses and reactive oxygen species generation in mice exposed to intermittent hypoxia. *The Journal of Physiology*, 577(2), pp.705–716.
- Peppard, P.E. et al., 2000. Prospective Study of the Association Between Sleep-Disordered. *The New England Journal of Medicine*, 342(19), pp.1378–1384.
- Pescador, N. et al., 2005. Identification of a functional hypoxia-responsive element that regulates the expression of the egl nine homologue 3 ( egl n3/phd3 ) gene. *Biochemical Journal*, 390(1), pp.189–197.
- Pillai, S.G. et al., 2009. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS genetics*, 5(3), p.e1000429.

- Prabhakar, N.R. et al., 2007. Systemic, cellular and molecular analysis of chemoreflex-mediated sympathoexcitation by chronic intermittent hypoxia. *Experimental physiology*, 92(1), pp.39–44.
- Prabhakar, N.R. & Semenza, G.L., 2012. Adaptive and Maladaptive Cardiorespiratory Responses to Continuous and Intermittent Hypoxia Mediated by Hypoxia-Inducible Factors 1 and 2. *Physiological Reviews*, 92(3), pp.967–1003.
- Price, S.J., Greaves, D.R. & Watkins, H., 2001. Identification of novel, functional genetic variants in the human matrix metalloproteinase-2 gene. Role of Sp1 in allele-specific transcriptional regulation. *Journal of Biological Chemistry*, 276(10), pp.7549–7558.
- Proceedings of the 4th World Symposium on Pulmonary Hypertension, February 2008, Dana Point, California, U., 2009. Proceedings of the 4th World Symposium on Pulmonary Hypertension, February 2008, Dana Point, California, USA. *J Am Coll Cardiol.*, 54(1), pp.S1–117.
- Pyzocha Neena et al., 2014. *Gene Correction* F. Storici, ed., Totowa, NJ: Humana Press.
- Qiu, P. et al., 2004. Mutation detection using Surveyor<sup>TM</sup> nuclease. *BioTechniques*, 36(4), pp.702–707.
- Quanjer, P. et al., 1993. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J Suppl.*, 16, pp.5–40.
- Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), pp.2281–2308.
- Reddy, T.E. et al., 2012. The effects of genome sequence on differential allelic transcription factor occupancy and gene expression. *Genome Research*, 22(5), pp.860–869.
- Redfield, M.M. et al., 2003. Burden of Systolic and Diastolic Ventricular Dysfunction in the Community. *JAMA*, 289(2), pp.194–202.
- Redline, S. et al., 1995. The familial aggregation of obstructive sleep apnea. *Am J Respir Crit Care Med.*, 151(3 Pt 1), pp.682–687.
- Regan, E. a et al., 2011. Genetic Epidemiology of COPD (COPDGene) Study Design. *Epidemiology*, 7(1), pp.1–10.
- Reijnen, M.J. et al., 1993. Hemophilia B Leyden: substitution of thymine for guanine at position -21 results in a disruption of a hepatocyte nuclear factor 4 binding site in the factor IX promoter. *Blood*, 82(1), pp.151–158.
- Riha, R.L. et al., 2005. Tumour necrosis factor- $\alpha$  (-308) gene polymorphism in obstructive sleep apnoea-hypopnoea syndrome. *European Respiratory Journal*, 26(4), pp.673–678.
- Rincón, M. & Flavell, R. a, 1994. AP-1 transcriptional activity requires both T-cell receptor-mediated and co-stimulatory signals in primary T lymphocytes. *The EMBO Journal*, 13(18), pp.4370–4381.
- Rosenbloom, K.R. et al., 2013. ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research*, 41(D1), pp.D56–D63.

- Rutter, J.L. et al., 1998. A Single Nucleotide Polymorphism in the Matrix Metalloproteinase-1 Promoter Creates an Ets Binding Site and Augments Transcription Advances in Brief A Single Nucleotide Polymorphism in the Matrix Metalloproteinase-1 Promoter Creates an Ets Binding Site an. *Cancer Research*, (10), pp.5321–5325.
- Sampsonas, F. et al., 2010. Positive association between two polymorphic sites (+134 insA/delA and G198T) of the endothelin-1 gene and chronic obstructive pulmonary disease. A case-control study. *Respiratory Medicine*, 104(1), pp.114–120.
- Schadt, E.E. et al., 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature*, 422(6929), pp.297–302.
- Schödel, J. et al., 2011. High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. *Blood*, 117(23), pp.e207–e217.
- Schödel, J. et al., 2012. Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nature Genetics*, 44(4), pp.420–425.
- Schofield, C.J. & Ratcliffe, P.J., 2004. Oxygen sensing by HIF hydroxylases. *Nature Reviews Molecular Cell Biology*, 5(5), pp.343–354.
- Schorg, A. et al., 2015. Destruction of a distal hypoxia response element abolishes trans-activation of the PAG1 gene mediated by HIF-independent chromatin looping. *Nucleic Acids Research*, 43(12), pp.5810–5823.
- Semenza, G.L., 2004. Hydroxylation of HIF-1: Oxygen Sensing at the Molecular Level. *Physiology*, 19(4), pp.176–182.
- Semenza, G.L. et al., 1996. Hypoxia Response Elements in the Aldolase A , Enolase 1 , and Lactate Dehydrogenase A Gene Promoters Contain Essential Binding Sites for Hypoxia-inducible Factor 1 \*. *The Journal of Biological Chemistry*, 271(51), pp.32529–32537.
- Semenza, G.L., 2003. Targeting HIF-1 for cancer therapy. *Nature Reviews Cancer*, 3(10), pp.721–732.
- Semenza, G.L. et al., 1994. Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *Journal of Biological Chemistry*, 269, pp.23757–23763.
- Semenza, G.L. & Wang, G.L., 1992. A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. *Molecular and Cellular Biology*, 12(12), pp.5447–5454.
- Serra, a. et al., 2002. Serotonin and serotonin receptor expression in the aorta of carotid intact and denervated newborns. *Respiratory Physiology and Neurobiology*, 132(3), pp.253–264.
- Shahar, E. et al., 2001. Sleep-disordered breathing and cardiovascular disease: cross- sectional results of the Sleep Heart Health Study. *Am.J.Respir.Crit.Care.Med*, 163(1), pp.19–25.
- Silverman, E.K. et al., 2011. Opportunities and challenges in the genetics of COPD 2010: an International COPD Genetics Conference report. *Copd*, 8(2), pp.121–135.

- Smemo, S. et al., 2014. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, 507(7492), pp.371–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24646999>.
- Soriano, J.B. & Lamprecht, B., 2012. Chronic Obstructive Pulmonary Disease: a worldwide problem. *Medical Clinics of North America*, 96(4), pp.671–680.
- Spielberger, C., Gorsuch, R. & Lushene, R., 1970. STAI, Manual for the State-Trait Anxiety Inventory (Self-evaluation Questionnaire). Palo Alto,. *Consulting Psychologists Press*.
- Stoller, J. & Aboussouan, L., 2005. Alpha1-antitrypsin deficiency. *Lancet*, 365(9478), pp.2225–36.
- Takeyama, K. et al., 2001. Activation of epidermal growth factor receptors is responsible for mucin synthesis induced by cigarette smoke. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 280(1), pp.L165–L172.
- Takeyama, K. et al., 1999. Epidermal growth factor system regulates mucin production in airways. *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), pp.3081–3086.
- The 1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. *Nature*, 467, pp.1061–1073.
- The GTEx Consortium, ., 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 348(6235), pp.648–60.
- The Joint European Society of Cardiology American College of Cardiology Committee, 2000. Myocardial infarction redefined—A consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the Redefinition of Myocardial Infarction. *European Heart Journal*, 21(18), pp.1502–1513.
- Thompson, A. & Elks, P., 2014. Hypoxia-inducible factor 2 $\alpha$  regulates key neutrophil functions in humans, mice, and zebrafish. *Blood*, 123(3), pp.366–376.
- Tian, H., McKnight, S.L. & Russell, D.W., 1997. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes & Development*, 11(1), pp.72–82.
- Torbicki, A. et al., 2008. Guidelines on the diagnosis and management of acute pulmonary embolism: The Task Force for the Diagnosis and Management of Acute Pulmonary Embolism of the European Society of Cardiology (ESC). *European Heart Journal*, 29(18), pp.2276–2315.
- Villar, D. et al., 2012. Cooperativity of Stress-Responsive Transcription Factors in Core Hypoxia-Inducible Factor Binding Regions J. Mata, ed. *PLoS ONE*, 7(9), p.e45708.
- Visel, A., Rubin, E.M. & Pennacchio, L. a, 2009. Genomic views of distant-acting enhancers. *Nature*, 461(7261), pp.199–205.
- Wallace, C. et al., 2006. Linkage analysis using co-phenotypes in the BRIGHT study reveals novel potential susceptibility loci for hypertension. *American Journal of Human Genetics*, 79(2), pp.323–331.



- Wanger, J. et al., 2005. Standardisation of the measurement of lung volumes. *European Respiratory Journal*, 26(3), pp.511–522.
- Wasserman, W.W. & Sandelin, A., 2004. Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics*, 5(4), pp.276–287.
- Wenger, R.H., Stiehl, D.P. & Camenisch, G., 2005. Integration of Oxygen Signaling at the Consensus HRE. *STKE*, (306).
- Wilk, J.B. et al., 2009. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS genetics*, 5(3), p.e1000429.
- Xia, X. et al., 2009. Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis. *Proceedings of the National Academy of Sciences of the United States of America*, 106(11), pp.4260–4265.
- Xia, X. & Kung, A.L., 2009. Preferential binding of HIF-1 to transcriptionally active loci determines cell-type specific response to hypoxia. *Genome Biology*, 10(10), p.R113.
- Yang, X., Guo, M. & Wan, Y.-J.Y., 2010. Deregulation of Growth Factor, Circadian Clock, and Cell Cycle Signaling in Regenerating Hepatocyte RXR $\alpha$ -Deficient Mouse Livers. *The American Journal of Pathology*, 176(2), pp.733–743.
- Yoo, S. et al., 2015. Integrative Analysis of DNA Methylation and Gene Expression Data Identifies EPAS1 as a Key Regulator of COPD. *PLoS Genetics*, 11(1), p.e1004898.
- Young, T. et al., 1993. The occurrence of sleep-disordered breathing among middle-aged adults. *The New England Journal of Medicine*, 328(17), pp.1230–5.
- Yu, C. et al., 2015. Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells. *Cell Stem Cell*, 16(2), pp.142–147.
- Yuan, G. et al., 2008. Induction of HIF-1 $\alpha$  expression by intermittent hypoxia: Involvement of NADPH oxidase, Ca<sup>2+</sup> signaling, prolyl hydroxylases, and mTOR. *Journal of Cellular Physiology*, 217(3), pp.674–685.
- Zhang, F., Wen, Y. & Guo, X., 2014. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Human Molecular Genetics*, 23(R1), pp.R40–R46.
- Zhang, Z. & Gerstein, M., 2003. Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Research*, 31(18), pp.5338–5348.
- Zhao, Y. et al., 2013. Association between 5-HT<sub>2A</sub> receptor polymorphisms and risk of obstructive sleep apnea and hypopnea syndrome: A systematic review and meta-analysis. *Gene*, 530(2), pp.287–294.
- Zhou, S. et al., 2015. A Single Nucleotide Polymorphism in 3' Untranslated Region of Epithelial Growth Factor Receptor Confers Risk for Pulmonary Hypertension in Chronic Obstructive Pulmonary Disease. *Cellular Physiology and Biochemistry*, 36(1), pp.166–178.

Ziółkowska-Suchanek, I. et al., 2015. Susceptibility loci in lung cancer and COPD: association of IREB2 and FAM13A with pulmonary diseases. *Scientific Reports*, 5(April), p.13502.

